BACTERIAL TRANSCRIPTIONAL RESPONSES TO CHEMICAL CHALLENGE

INVESTIGATING THE TRANSCRIPTIONAL RESPONSES OF ESCHERICHIA COLI UNDER CHEMICAL CHALLENGE

BY BING YA GUO, B.H.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree
Master of Science

MASTER OF SCIENCE (2020)
Department of Biochemistry and Biomedical Sciences

McMaster University
Hamilton, Ontario

TITLE: Investigating the transcriptional responses of *Escherichia coli* under chemical challenge

AUTHOR: Bing Ya Guo, B.H.Sc. (McMaster University)

SUPERVISOR: Professor E.D. Brown

NUMBER OF PAGES: xi, 68

LAY ABSTRACT

In the current era of antibiotic resistance, there is an urgent need to discover novel chemicals that can be used to treat bacterial infections. The modern antibiotic drug discovery pipeline allows large-scale chemical screening to test for the ability of a potential drug to kill off or stop the growth of bacteria. However, a common bottleneck associated with this lengthy and complex process is the difficulty in understanding how these new drugs are fighting off bacterial infections; this is called the mechanism of action (MOA) of a chemical. This is extremely important as it can help researchers gain a better understanding of the likelihood that bacteria will develop resistance to the drug rapidly. It can also reveal information on potential toxicity of the drug to humans. Ultimately, exploring the specific MOA of a new drug candidate can help researchers prioritize which drug to take to clinical trials. In this study, we investigate how bacteria respond to antibiotic treatments by looking at how they alter their genome to adapt to this chemical stress. We observed that similarities and differences in the changes to their genetic blueprint can help researcher categorize compounds with similar MOAs into the same group. Using this information, we have built a machine learning model that can help predict the MOA of a compound. This presents an exciting opportunity for researcher to speed up their discovery pipeline and help bring new antibiotics to the market.

ABSTRACT

Bacteria live in diverse and dynamic environments that necessitate adaptation to external stimuli. Here, we probe *Escherichia coli* K-12 with a wide array of chemical stressors spanning several drug classes, and gauge the transcriptional responses using a promoter-GFP fusion library. Assayed with PFIboxes, the output fluorescence images are temporally resolved and data rich. When quantified as gene expression, these transcriptional responses are seemingly unique to each molecule tested, yet clear differences exist between drug classes. Promoters showing large magnitudes of differential regulation in selective conditions, such as DNA damage, oxidative stress, and cell wall stress, can be used as diagnostic reporters for primary screening assays. The transcriptional signatures generated by these experiments were used to train a 10-layer convolutional neural network in Keras for mechanism of action (MOA) predictions. This model was used to predict the mechanism of action of cefmetazole, polymyxin B, as well as cinoxacin, a compound excluded from the training set. The model predicted the identity of cefmetazole and polymyxin B with 95-100% accuracy. Cinoxacin was predicted to be enoxacin, another fluoroquinolone antibiotic, with ~80% confidence, illustrating the power of prediction MOA of unknown molecules with a large training dataset. This deep learning model predicted the MOA of an unknown compound, MAC168425, as trimethoprim. Further characterization of the compound suggests that its inhibitory activity is involved in folate-binding and utilization, and glycine cleavage. In all, this work illustrates that microbial reporter arrays generate unique patterns which can be used to make hypotheses on the MOA of unknown molecules.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Eric Brown, for giving me the opportunity to conduct my undergraduate thesis in his lab three years ago. It really ignited my interest and passion for research and inspired me to pursue graduate research in the lab. The past few of years have been exciting and challenging in the best way. Thank you for providing such a creative and collaborative environment to train in. Thank you to my supervisory committee members, Dr. Mike Surette and Dr. Marie Elliot, for all of your guidance and feedback.

I would also like to thank Shawn for his tireless mentorship. Thank you for your creativity, your continuous guidance, and for teaching me how to start thinking like a scientist. Please do not ever change. Thank you to Mike for being a great teacher who is always open to discussions, and for guiding me in the right direction. Thank you to Ken for getting through the thick and thin with me. Thank you to Brent, Sara, Kristina, Maddy, and Garima for your endless and selfless support. You guys always know what to say. Thank you to Jodi for always taking care of us. To all of the Brownies, I like you all more than I care to admit.

Finally, I would like to thank my parents for providing me the opportunity to live a life where I am free to pursue what I enjoy. Thank you for carrying so many burdens for me and caring for me unconditionally.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASKA	a complete set of *E. coli* K-12 ORF archive
BARseq	barcode sequencing
bp	basepair
cDNA	complementary deoxyribonucleic acid
DNA	deoxyribonucleic acid
DMSP	dimethylsulfoniopropionate
EDTA	ethylenediaminetetraacetic acid
GO	gene ontology
GFP	green fluorescent protein
IPTG	isopropyl β-D-1-thiogalactopyranoside
IS5	insertion sequence 5
Kan	kanamycin
LB	lysogeny broth
LDA	linear discriminant analysis
LPS	lipopolysaccharide
MIC	minimum inhibitory concentration
MOA	mechanism of action
MOPS	3-(N-morpholino)propanesulfonic acid
NADPH	nicotinamide adenine dinucleotide phosphate
OD_{600}	optical density measured at 600 nm
PFIbox	printed fluorescence imaging box
RNA	ribonucleic acid
RNAseq	ribonucleic acid sequencing
rpm	revolutions per minute
RT-qPCR	quantitative reverse transcription PCR
SDS	Sodium dodecyl sulfate
TE	Tris ethylenediaminetetraacetic acid
TSS	transformation and storage solution

DECLARATION OF ACADEMIC ACHIEVEMENT

Shawn French performed the linear discriminant analysis and generated the deep learning model for predicting compound MOA. He created the corresponding figures (Figure 4 and 9) and co-wrote the methods and results sections for these analyses.

Jean-Philippe Côté generated the MAC168425 resistant mutants.

CHAPTER 1: INTRODUCTION

In the current era of antibiotic resistance, drug target identification and mechanism of action (MOA) are both highly sought-after but elusive components in the modern drug discovery pipeline[1]. Recent high-throughput chemical screens tend to assess for antibiotic activity within a chemical library in a relevant condition. In particular, cell-based screening approaches are phenotypic based and can identify chemical perturbants of biology without being limited to a single target. However, investigating the MOA of newly discovered bioactive compounds is a lengthy and complex process, as compounds found in phenotype-based screens can affect multiple cellular targets. Inhibition of these targets commonly results in multifaceted downstream effects that extend beyond simple enzyme inhibition[1,2]. While direct protein targets have been identified for most conventional antibiotics, indirect and secondary effects to these antibiotics are often poorly characterized[2]. Elucidating these complex stress responses can help to define drug tolerance and resistance[3]. Furthermore, rapid identification of drug MOA can help researchers prioritize lead compounds from any chemical screening in a more meaningful way. Ultimately, there is a need to explore how bacteria respond to antibiotic stress on a systems level.

1.2 Bacteria respond to external stimuli through various stress responses

Bacteria exist in unpredictable environments and are often subjected to a variety of stressors and must sense and respond through adaptation[2]. These include antimicrobial agents, hydrogen ions, or nutrient deprivation, amongst others[2]. In nutrient limited

conditions, for example, the stringent response is often initiated[4]. This is a global physiological response, arising from stresses such as restrictions in amino acids, carbon sources, iron, and phosphates[4]. This response is mediated by the alarmone (p)ppGpp, whose levels are governed by the widely conserved RelA/SpoT homolog enzymes depending on nutrient availability[5,6]. (p)ppGpp acts as a global transcriptional regulator by modulating RNA polymerase activity to help divert cell resources from protein synthesis to activating metabolic biosynthesis processes[5,7]. Another example is the envelope stress response, wherein transcriptional responses are activated by perturbation of different components of the cell envelope[2]. For instance, the Psp response helps to stabilize the proton motive force of the cell when the inner membrane integrity is impaired[8,9]. Additionally, a variety of outer-membrane related perturbations, such as mutations in the periplasmic chaperones[10] and in genes that alter lipopolysaccharide (LPS)[11], have been described to induce the σ^E – dependent extracytoplasmic stress response. σ^E is known to regulate over 60 transcriptional units in *E. coli*, most of which are involved in the biosynthesis, folding, and homeostasis of outer membrane proteins and LPS, and can help target mis-folded proteins for degradation[2,12].

1.3 Antibiotic challenge leads to differential gene expression

Sub-inhibitory concentrations of antibiotics can also trigger diverse changes in gene expression. These transcriptional responses result in changes in the regulation of steps directly affected by target inhibition. For example, the fluoroquinolone group of antibiotics[13] are potent, bactericidal antibiotics that inhibit DNA gyrase and

topoisomerase[14]. This inhibition eventually results in the formation of double-stranded DNA breaks, stalled DNA replication forks and, ultimately, cell death[13,14]. At sub-inhibitory concentrations, they are potent inducers of the SOS response, a global regulatory network in response to DNA damage in which the cell cycle is arrested and DNA repair is initiated[15]. The same holds true for non-quinolone-based DNA damaging agents, such as cisplatin and mitomycin C, as well as folate biosynthesis inhibitors, such as a trimethoprim, all of which trigger the SOS response[16]. The SOS response leads to the induction of a cascade of over 50+ genes involved in high fidelity DNA repair (e.g., *polB*[17]), inhibition of cell division (e.g., *sulA*[18]), and low fidelity DNA tolerance repair pathway (e.g., *umuC*[19])[20]. Similarly, exposure to fatty acid biosynthesis inhibitors, such as triclosan and cerulenin, which targets the FabI[21] and FabB[22] enzymes respectively, have been shown to cause transcriptional induction of the *fabI, fabF, fabA,* and *fabB* genes[23].

Beyond the direct transcriptional response to the inhibition of the primary target, expression alterations are often complicated by secondary target inhibition or indirect effects. For example, Mitosch et al. demonstrate that at sublethal concentrations of trimethoprim, *Escherichia coli* will initiate rapid acid stress response by upregulating *gadBC,* a major acid stress operon[24]. Cells with higher *gadBC* expression following trimethoprim treatment are able to maintain higher intracellular pH and survive subsequent acid challenge[24]. These indirect or secondary effects can provide a better understanding of the downstream effects of antibiotics or the compensatory mechanisms that may arise due to changes in the environment. Ultimately, each class of antibiotics may induce a distinct set of promoters that are unique to that class and is reflective of the

inhibition of its target. However, signatures of individual compounds may also include transcriptional alterations on secondary targets or downstream effects, which may not be conserved within the same class of antibiotics nor unique to a single MOA class.

1.4 Genome-wide studies reveal diverse changes in patterns of transcriptional regulation

Genome-wide queries into antibiotic responses have been undertaken using DNA microarray[25], proteomic investigations[26], or a global transcriptional method[27]. Namely, Goh et al. measured bacterial transcription patterns using promoter-*lux* reporter constructs in a 6,500-clone *Salmonella typhimurium* library under erythromycin and rifampicin stress[27]. In this study, approximately 5% of the promoters were found to be modulated in the presence of sub-inhibitory concentrations of antibiotics, and these active promoters respond to varying extents depending on the antibiotic and drug concentration being used[27]. These promoters help transcribe genes of various function, such as transport, virulence, and DNA repair.

Furthermore, Utaida et al. conducted a genome-wide transcriptional profiling experiment to further our understanding on the molecular events occurring upon challenge with the cell wall active antibiotics oxacillin, bacitracin, and D-cycloserine[28]. More than a hundred genes were commonly regulated by these three antibiotics; they belong to various functional categories, such as cell-envelope biogenesis, DNA replication, amino acid transport and metabolism[28]. More interestingly, there are more than 300 differentially expressed genes that are unique to just one of these antibiotics[28].

These findings suggest that antibiotics with different chemical structures and enzyme targets will lead to diverse changes in patterns of bacterial gene expression, even if they belong to the same general MOA class[27]. An understanding of genome-wide gene expression fingerprints in response to compounds of known MOA may help infer the MOAs of novel uncharacterized chemicals.

1.5 Transcriptional signature can help classify antibiotics based on mechanisms of action

Leveraging gene expression profiling as a predictor of antibiotic MOA is not a novel approach and has been approached in several manners in the past. A group of researchers led by Dr. Hannes Loferer investigated the global transcript levels of a Gram-positive bacterium, *Bacillus subtilis,* in nutrient-rich media following treatment with 37 known antibiotics spanning 6 MOA classes using microarrays[23]. Predictions were made using a support vector machine, which is able to classify compounds based on differences between the MOA classes[23]. This model achieved high success rates (>80%) in MOA predictions in most of the antibiotic classes[23]. However, DNA damaging agents azaserine, doxorubicin, and hydrogen peroxide were categorized with the group of quinolone antibiotics despite not having the same enzyme target[23]. This result suggests that the predictor is not able to differentiate MOA within a broad group of chemicals that cause DNA stress[23]. Furthermore, the small sample size of merely 2-4 compounds under certain classes, such as folate and fatty acid biosynthesis compounds, resulted in false predictions for many known and test compounds[23].

Further, these data were then used to identify marker genes that are indicative of certain compound classes[29]. Differentiating reporter strains were identified for inhibitors of protein biosynthesis (*yrzI*), fatty acid biosynthesis (*fabHB* and *glpD*), cell wall biosynthesis (*ypbG*), as well as quinolones (*dinB*) and glycopeptide antibiotics (*ytrA* and *ywoB*)[29]. However, it was difficult to determine the reporter strains for classes of protein and cell wall biosynthesis, as no promoter fulfilled the criteria of their filter[29]. As such, these reporters led to misclassification of their test compounds[29]. Together, these findings indicate that it would be valuable to increase the number of compounds in the database to improve the accuracy of the prediction tool and the specificity of fingerprint reporter strains.

Transcriptional assays using next-generation sequencing, such as RNA-seq, have also been used to speculate on MOA for unknown antibiotics. A group led by Dr. Karen Nelson conducted RNAseq analyses of bacterial transcriptomics under challenge from 37 antibiotics spanning 6 different MOA classes[30]. Through supervised clustering, two cell wall (*wca* and *ent*/*fep* specific for fosfomycin) and one fatty acid synthesis (*fabI*/*fabB*) expression signatures were discovered based on the magnitude of the transcriptional response[30]. However, these diagnostic genes were often modestly regulated (3-8-fold) by many small molecules from other MOA classes[30], suggesting that diagnoses from these genes may be prone to false positives. More than 400 diagnostic genes were selected based on those that are most significantly regulated for each compound[30]. The MOA of the protein synthesis, fatty acid synthesis, and cell wall biosynthesis inhibitors were predicted with 100% accuracy, while the DNA synthesis and cell membrane groups

achieved around 50% accuracy[30]. This suggests that investigations of transcriptomics can yield diagnostic responses to antibiotic treatment. A larger dataset performed in nutrient limited medium has the potential to reveal diagnostic reporter genes for inhibitors of metabolic pathways and uncover additional diagnostic genes for other biological processes.

1.6 Research Objectives

The goal of this research project is to investigate the transcriptional responses of *E. coli* when subjected to chemical challenge in nutrient limited conditions. Zaslaver *et al.* have created a 1,820 promoter-GFP fusion library to measure transcriptional activity in *E. coli*[31]. Conventional plate reader hardware and consumable options do not allow for screening beyond 1,536-density microwell plates and upscaling this approach to a full screening platform allows for high throughput acquisition of time-course global gene expression data in *E. coli*[31]. Using solid media arrays and custom hardware (PFIboxes), it is possible to acquire images of fluorescence phenotypes with high temporal resolution in a simple and inexpensive manner[32]. This utilizes the transcriptional responses of *E. coli* to drug stress, measuring global promoter activity in *E. coli* by means of promoter-reporter fusion constructs. PFIbox screening pipelines produce a wealth of multidimensional data, in 6,144-density, for every screened chemical.

I hypothesize that antibiotics of the same class will induce similar stress responses in bacteria, and that each class will result in a unique set of gene expression profile that can help predict the mechanism of action of an unknown molecule. My objectives are to:

1) investigate the basic biology behind *E. coli*'s transcriptional network under antibiotic challenge

2) build a comprehensive training set of transcriptional profiles for known antibiotics to help predict chemical mechanisms of action

3) characterize the MOA of MAC168425, a compound with an unknown mechanism of action

CHAPTER 2: MATERIALS AND METHODS

2.1 Reagents

All chemicals and reagents were purchased from Sigma Aldrich unless otherwise specified. MAC168425 was purchased from Accela Chembio. Antibiotics were added to the media as needed for selection with final concentrations as follows: 25 µg/ml kanamycin, 20 µg/ml chloramphenicol, and 50 µg/mL ampicillin if needed, unless otherwise stated. See Table 1 for a list of chemicals and concentrations used in this study.

Table 1. List of chemicals screened with the GFP-promoter library in this study.

Chemical	MOA Class Classification	Solid MIC (µg/mL)	Screening Concentrations (µg/mL)
Ampicillin	Cell Wall	16	4, 2
Apramycin	Protein	2	1, 0.25
A22	Other	>256	128, 64, 32
Azidothymidine	DNA	>256	256
Azithromycin	Protein	128	64, 32, 16
CHIR-090	Membrane	0.125	0.031, 0.016
Cefadroxil	Cell Wall	1	0.25
Cefazolin	Cell Wall	1	0.25
Cefmetazole	Cell Wall	2	1, 0.5
Cefoxime	Cell Wall	1	0.25
Cefoxitin	Cell Wall	4	1
Ceftazidime	Cell Wall	2	0.25
Cerulenin	Fatty Acid	128	32, 16
Chloramphenicol	Protein	16	8, 4, 2
Cinoxacin	DNA	32	16, 8, 4
Ciprofloxacin	DNA	0.25	0.0625, 0.03
Colistin	Membrane	16	8
D-cycloserine	Cell Wall	0.5	0.25, 0.125
Dapsone	Folate	256	64, 32
Dirithromycin	Protein	>256	256, 128
Doxycycline	Protein	16	8, 4, 2
EDTA	Membrane	>256	256
Enoxacin	DNA	8	4, 2, 1

Erythromycin	Protein	>256	256, 64
5-fluoroanthranilic acid	Nutrient	4	2, 1, 0.5
5-fluorouracil	Nutrient	4	2
5-methyltryptophan	Nutrient	16	2
Fosfomycin	Cell Wall	1	0.5
Furazolidone	DNA	8	4, 1
Fusidic acid	Protein	>256	128
Gentamicin	Protein	0.25	0.031
Glyphosate	Nutrient	>256	64
Imipenem	Protein	1	0.5, 0.25, 0.125
L-norleucine	Nutrient	128	64, 32
L-3-thienylalanine	Nutrient	128	16
Lincomycin	Protein	>256	128
Linezolid	Protein	>256	256, 64
MAC13772	Nutrient	>256	256
MAC168425	Other	128	64, 32, 16
MAC173979	Nutrient	>256	64
MAC872	Other	128	32
Mecillinam	Other	256	128, 64
Meropenem	Cell Wall	2	0.5, 0.25
Metronidazole	DNA	>256	256, 64
Minocycline	Protein	4	1, 0.5
Mitomycin C	DNA	2	0.25
Nalidixic acid	DNA	32	16, 8
Neomycin	Protein	8	4, 1
Norfloxacin	DNA	2	1
Novobiocin	DNA	>256	128, 64
Paraquat	Other	16	4, 2
Penicillin G	Cell Wall	128	32, 16
Pentamidine	Membrane	>256	64
PF 5081090	Membrane	0.125	0.0625, 0.0156
Polymyxin B	Membrane	2	1, 0.5, 0.25
Polymyxin B nonapeptide	Membrane	>256	50, 25, 12.5
Rifampicin	mRNA	32	8
6-diazo-5-oxo-L-norleucine	Nutrient	0.063	0.031, 0.016, 0.008
6-mercaptopurine	Nutrient	>256	256
6-aminoindole	Nutrient	>256	128, 64
Sodium bicarbonate	Other	>50 mM	25 mM, 12.5 mM
Spectinomycin	Protein	64	16

SPR741	Membrane	32	8
Streptomycin	Protein	2	1, 0.5, 0.25
Sulfadimethoxine	Folate	256	128, 64, 32
Sulfamethoxazine	Folate	>256	256, 128, 64
Sulfathiazole	Folate	>256	256, 128, 64
Sulfamethizole	Folate	256	64
Sulfamethoxazole	Folate	256	128, 64, 32
Sulfisoxazole	Folate	>256	256, 128, 64
Tetracycline	Protein	16	8, 4, 2
Triclosan	Fatty Acid	0.5	0.25, 0.125, 0.0625
Trimethoprim	Folate	4	2, 0.25
2 2'-bipyridyl	Other	64	16, 8

2.2 Strain library preparation and growth conditions

For a detailed description of the library preparation and growth conditions, refer to the protocol described by French *et al*[32,33]. The GFP promoter collection[31] was grown from frozen stocks at 384 density onto Singer PlusPlates (Singer Instruments, UK) filled with 25 mL of Lysogeny Broth (LB) agar (1.5%) medium supplemented with 25 µg/mL of kanamycin. These plates were grown at 37 °C for 18 hours, then upscaled to 1536 density onto MOPS minimal medium supplemented with 0.4% glucose (Teknova, US) and 25 µg/mL of kanamycin using the Singer Rotor (Singer Instruments, UK). These plates were grown at 37 °C for 24 hours.

2.3 Solid MIC determination

The minimum inhibitory concentration (MIC) for each chemical in solid media was determined as described by French *et al*[34]. The liquid MICs were established for each compound to provide a reference point for the concentrations to be used in the solid

potency assay. Solid media plates (Singer Instruments, UK) were filled with 25 mL of 2% agarose to prepare a mold for the media plugs. Empty plugs were filled with increasing concentrations of the test compound mixed with media until volume is leveled with the agarose bed. The agarose bed was removed and the media plugs were inoculated with *E. coli* K-12 MG1655[31], using the same Singer Rotor settings as those used for the screening assay plates. These plates were grown at 37°C for 24 hours, then MIC determined from the plugs containing no colonies.

2.4 Gene expression assay

For a detailed description of the experimental setup and analysis, refer to the protocol described by French *et al*[32,33]. The Alon collection was probed against a panel of antibiotics at sub-inhibitory concentrations (1/2-1/8x MIC). MOPS minimal medium supplemented with 0.4% glucose, 25 µg/mL of kanamycin, and the screening antibiotic was poured at 25 mL per Singer PlusPlate as per French *et al*[33]. Plates were poured on the day of the experiment and inoculated to 6144 density from prepared master reporter library plates. Plates were placed face down in PFIboxes[32] and incubated at 37°C for 24 hours, imaging every 5 minutes.

2.5 Data preparations and visualization

Fluorescent images were capture using PFIboxes[32] and analyzed using ImageJ[35] to extract fluorescent intensity values for each colony. Cumulative fluorescence was calculated for 24 hours of growth on MOPS minimal medium with sub-inhibitory concentrations for

each drug screened. This provided unique overall fingerprints of promoter activity across the duration of the experiment, for each drug tested. Edge effects were normalized using a two-pass row and column normalization system[34]. Fingerprints were compiled as data frames for use in downstream analyses. In the linear discriminant analysis, known chemical classes were used as the groupings. These groupings were visualized in component space, where the first 3 discriminants comprised about 75% of the variances observed. Hierarchal clustering and heat map visualization were performed using the heatmap.2 function from the gplots package in R. Gene ontology (GO) term enrichments were performed through EcoCyc[36] to study differentially regulated biological processes.

2.6 Deep learning model and predictions

The deep learning model utilized the fluorescence patterns at each individual timepoint, for each drug tested. Using these data, a deep learning network was built using the Keras package in R, with Tensorflow as the backend. A 10-layer model was constructed consisting of: 2D convolution (64 filter, 5x5 kernel, rectified linear unit (relu) activation), 0.25 dropout, 2D convolution (128 filter, 3x3 kernel, relu activation), 0.25 dropout, 2D pooling (pool size 4), 0.25 dropout, flattened layer, densely connected layer (50 unit, relu activation), 0.25 dropout, densely connected layer (softmax activation). The input layer had 6,144 neurons, the number of colonies in a 64 x 96 array, and the network was compiled with the Adam optimizer with a binary cross-entropy loss function. Accuracy was visualized alongside loss while the model was compiled, and both measures levelled out after 10 epochs. Internal validations were performed with an 80/20 random split of

the data, with 80% of the data acting as the training set and the remaining 20% as the validation set. For test purposes, sub-inhibitory concentrations of the antibiotics were used, to test which compound it best matched with in the training set. These compounds were not included in the original training set; they were assayed on a different day with a different frozen stock of the library and a different batch of media.

2.7 Bacterial strains and culture conditions for liquid assays

Single gene deletion strains used in this study were cultured from the Keio collection[37]. Protein overexpression strains and plasmids were cultured and isolated from the *E. coli* ASKA library[38]. Refer to Table 2 for a list of strains and plasmids used in this study.

Table 2. Bacterial strains and plasmids used in this study. All strains used in this study were *E. coli* strains.

Strain/Plasmid	Genotype Description	Source
Strains		
MG1655	wildtype	MG Surette Lab
BW25113	wildtype	Baba et al.[37]
	Δ*csrD::Kan*	Baba et al.[37]
	Δ*cycA::Kan*	Baba et al.[37]
	Δ*gcvA::Kan*	Baba et al.[37]
	Δ*gcvB::Kan*	ED Brown Lab
	Δ*gcvH::Kan*	Baba et al.[37]
	Δ*gcvP::Kan*	Baba et al.[37]
	Δ*gcvT::Kan*	Baba et al.[37]
	Δ*lrp::Kan*	Baba et al.[37]
	Δ*nudB::Kan*	Baba et al.[37]
	Δ*pabC::Kan*	Baba et al.[37]
	Δ*sstT::Kan*	Baba et al.[37]
	ΔygfA*::Kan*	Baba et al.[37]

	Δ*ygfZ::Kan*	Baba et al.[37]
	Δ*yhdH::Kan*	Baba et al.[37]
AG1	pCA24N	Kitagawa et al.[38]
	pCA24N-*folA*	Kitagawa et al.[38]
	pCA24N-*yhdH*	Kitagawa et al.[38]
Plasmids		
pSim6	pSC101 with λ bacteriophage red genes under native control; Amp^R	Datta et al.[39]
pCA24N	IPTG inducible overexpression (empty vector) by lacI; CM^R	Kitagawa et al.[38]
pCA24N-*folA*	IPTG inducible overexpression of *folA* by lacI; CM^R	Kitagawa et al.[38]
pCA24N-*yhdH*	IPTG inducible overexpression of *yhdH* by lacI; CM^R	Kitagawa et al.[38]

Overnight cultures of *E. coli* K-12 wildtype strain BW25113 were grown from a single colony in MOPS minimal medium (Teknova, US) supplemented with 0.4% glucose at 37 °C with shaking at 250 rpm. Saturated overnight cultures were diluted 1/50 into fresh MOPS medium with 0.4% glucose and grown at 37 °C with shaking until the OD_{600} reached 0.5. Subcultures were diluted 1/1000 to a final inoculum of 10^5 CFU/mL into fresh medium for assays in 96 well microplates. For the pSIM6 habouring strain containing a temperature sensitive plasmid, cultures were incubated at 30 °C.

2.8 TSS competent *E. coli* preparation and plasmid transformation

Bacterial cultures were prepared as stated above. Cells were harvested by centrifugation at 4000 x g for 15 minutes at 4 °C. The pellet was resuspended in 100 µL of TSS (100 mg/mL PEG 8000, 5% (v/v) DMSO, 5% (v/v) of 1M $MgCl_2$, 85% LB medium). Each reaction contained 2 µL of the plasmid of interest. The culture was incubated on ice for 5

minutes. Each reaction was supplemented with 200 µL of LB medium and incubated at 37 °C with shaking at 250 rpm for 45 minutes. Cultures were plated onto appropriate antibiotic plates and incubated overnight at 37 °C for selection of single colonies harbouring the plasmid of interested.

2.9 Liquid MIC determination

Bacterial cultures were prepared as stated above. Diluted subcultures were mixed with 2-fold serial dilutions of test compounds to a final volume of 200 µL. OD_{600} readings were measured using the Tecan Infinite M1000 plate reader prior to incubation to account for background absorbance. Microplates were incubated in a stationary 37 °C incubator for 24 hours. The minimum inhibitory concentration for each compound was determined as the lowest concentration resulting in ≤ 10% of the growth of the untreated culture.

2.10 Checkerboard assays

Bacterial cultures were prepared as described above. The checkerboard broth microdilution assays were prepared in 8x8 or 8x12 matrices of differing combinations of chemical concentrations. All compounds were prepared by 2-fold serial dilutions and mixed with diluted subcultures to a final volume of 200 µL. OD_{600} readings were measured using the Tecan Infinite M1000 plate reader prior to incubation to account for background absorbance. Microplates were incubated in a stationary 37 °C incubator for 24 hours, after which the post-incubation OD_{600} readings were measured.

2.11 Spontaneous resistant mutant generation and characterization

Spontaneous resistant mutants were generated by plating the wildtype *E. coli* BW25113 strain on 128 µg/mL of MAC168425 for 48 hours at 37 °C and isolating single colonies. Individual colonies were passaged in Lysogeny Broth (LB) medium without drug selection to confirm resistance to MAC168425. Genomic DNA of these strains were isolated using the Invitrogen PureLink™ Genomic DNA kit. Whole genome sequencing was conducted using the Illumina NextSeq 550 platform. Subsequent variant calling analyses were performed using the breseq pipeline.

2.12 RNA extraction

Overnight cultures of *E. coli* BW25113 wildtype and the MAC168425 spontaneous resistant strains were used to prepare subcultures as described above. RNA stop solution (10% acid-phenol:chloroform in ethanol, v/v) was added (1/10) to the subcultures and placed on ice immediately for 5 minutes. Cells were harvested by centrifugation (4000 x g) for 20 minutes at 4 °C. Cells were resuspended with 800 µL of freshly prepared lysis solution (20 mM sodium acetate, 1 mM EDTA, 0.5% SDS), then mixed with hot (65 °C) acid phenol solution. Samples were incubated at 65 °C for a total of 12 minutes, with mixing every 1 minute. The aqueous phase was isolated and extracted once more with acid-phenol:chloroform (125:24:1, pH 4.5, Invitrogen Ambion) and twice with phenol:chloroform:isoamy alcohol (25:24:1, pH 6.6, Invitrogen Ambion). The aqueous phase of the last extraction was added to 800 µL of 100% Ethanol. Further RNA

purification and genomic DNA removal was performed using the Monarch Total RNA Miniprep Kit (New England Biolabs).

2.13 Reverse transcription quantitative polymerase chain reaction

Total RNA (2 µg) was reverse transcribed to cDNA with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). cDNA was diluted with TE buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA) to a final concentration of 50 ng/µL. Each RT-qPCR reaction contained 5 ng of cDNA, 500 nM of each forward and reverse primer, and 1x of the PowerUP SYBR Green Master Mix (Applied Biosystems) in a total volume of 20 µL. RT-qPCR was performed using the CFX96 Real-Time System (BioRad). The cycle conditions were 50 °C for 2 minutes, 95 °C for 2 minutes, and 40 cycles of 95 °C for 15 seconds, and 62 °C for 40 seconds. Relative transcript levels were calculated using the Pfaffl method[40], with the *E. coli* BW25113 16S rRNA, *rrsA*, as a reference control. Refer to Table 3 for the sequences of each RT-qPCR primer used in this study.

Table 3. List of oligonucleotide sequences for primers used in the RT-qPCR gene expression experiments of this study.

Primer	Sequence (5'-3')	Source
csrD-F	TCGGCCTCTGCAAGTTCAAT	This study
csrD-R	AGGTTACCGTTGAGTCGCTG	This study
yhdH-F	TGGGTTATCAGGTCGTTGCC	This study
yhdH-R	ACGGGATTCGGCAAACTCAT	This study
rrsA-F	TATCCTTTGTTGCCAGCGGT	This study
rrsA-R	CGCTTCTCTTTGTATGCGCC	This study

CHAPTER 3: RESULTS

3.1 Generating transcriptional response signatures for known antibiotics

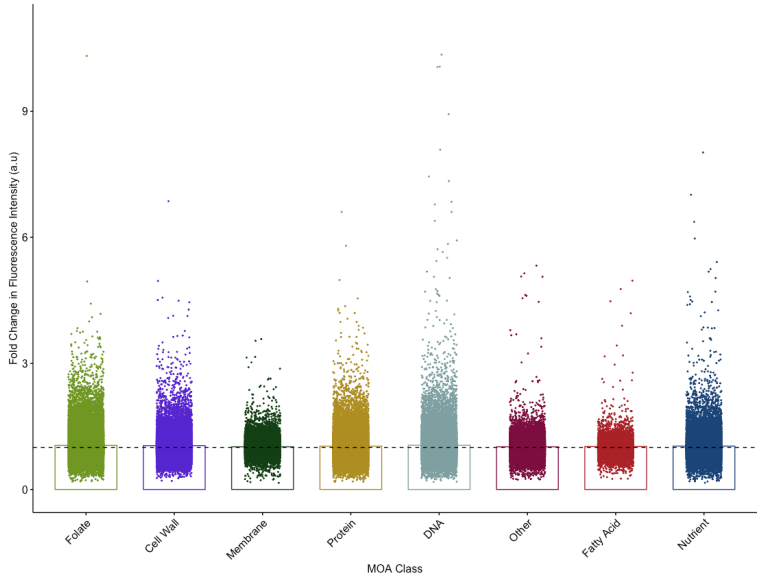
To study the transcriptional responses of *E. coli* to antibiotic stress, we used the GFP promoter-reporter fusion library created by Zaslaver et al.[31], which contains approximately 1,800 promoter-report fusion strains that allow for quantification of gene transcription, alongside the PFIboxes by French et al.[3,33] for high-throughput acquisition of time-course data. Since inhibitors of nutrient biosynthesis pathways are only active in nutrient limited media, the screen was conducted in MOPS minimal medium supplemented with 0.4% glucose as the carbon source. In brief, the promoter collection was arrayed in biological triplicate per treatment condition on a single solid medium plate containing over 6,000 colonies in total. These assay plates were then placed into a parallel assembly of 16 PFIboxes inside a single incubator, where GFP fluorescence images were acquired over time. An image analysis pipeline written by French et al.[32,33] was able to extract and provide fluorescence time-course data for every reporter strain in the library. These fluorescence data files were then compiled and organized into matrices of raw data. The edge effects were normalized using a method developed for high-density colony array[41], in which the colony fluorescence was divided by the interquartile means of the rows and columns across the plate. This method also standardizes fluorescence intensity values across plates. Technical and biological replicates of the same conditions showed a strong, linear correlation in terms of fluorescence intensity, indicating that the data was reproducible[32]. For detailed descriptions of the data acquisition and analysis pipeline, refer to the protocol described by French et al[33]. Ultimately, these genome-wide

transcriptional regulation studies were performed for a wide array of antibiotics spanning many MOA classes (Table 1). Antibiotics were divided into classes by their primary mechanism of action responsible for their inhibitory activity.

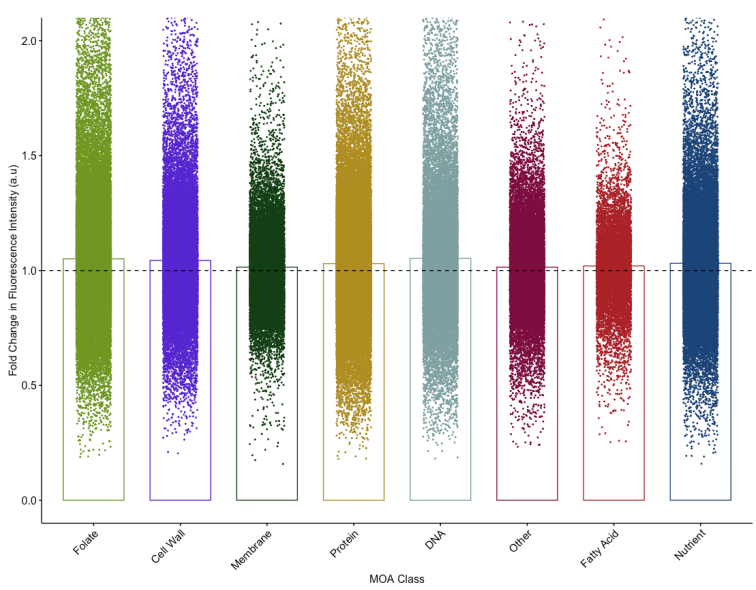
3.2 A summary of the spread of differentially regulated promoters in response to antibiotics from various MOA classes

To get an indication of the number of interactions between each promoter against every treatment condition, chemicals were categorized by their MOA class and visualized using a bar plot to showcase the spread of differentially regulated promoters (Figure 1A). Fold change in fluorescence intensity is calculated by dividing the drug treated condition by the untreated sample. The average fold change in fluorescence intensity is very close to 1 for all MOA classes (Figure 1B). However, many promoters were differentially regulated for each MOA class (Figure 1C). The number of upregulated and downregulated promoters were categorized by their magnitude of fold change. Here, we observed that challenge by some chemical classes, such as the DNA damaging class of antibiotics, results in many differentially regulated promoters that were induced to a higher magnitude relative to other groups. On the other hand, chemicals that belong to the fatty acid and membrane perturbing groups of antibiotics resulted in a smaller number of transcriptional responses. Transcriptional responses that showed a fold change of more than 2 and less than 0.5 were deemed to be differentially regulated in this screen. In summary, there are 3865 differential regulation phenotypes across all conditions.

A



B



C

Fold Change	Folate	Cell Wall	Membrane	Protein	DNA	Other	Fatty Acid	Nutrient
< 0.5	346	206	69	514	340	153	51	285
> 2	627	316	35	394	552	55	24	263
> 3	59	39	5	46	95	15	8	48
> 4	5	10	0	10	37	8	4	19
> 5	1	1	0	2	21	4	0	8
> 6	1	1	0	1	12	0	0	3

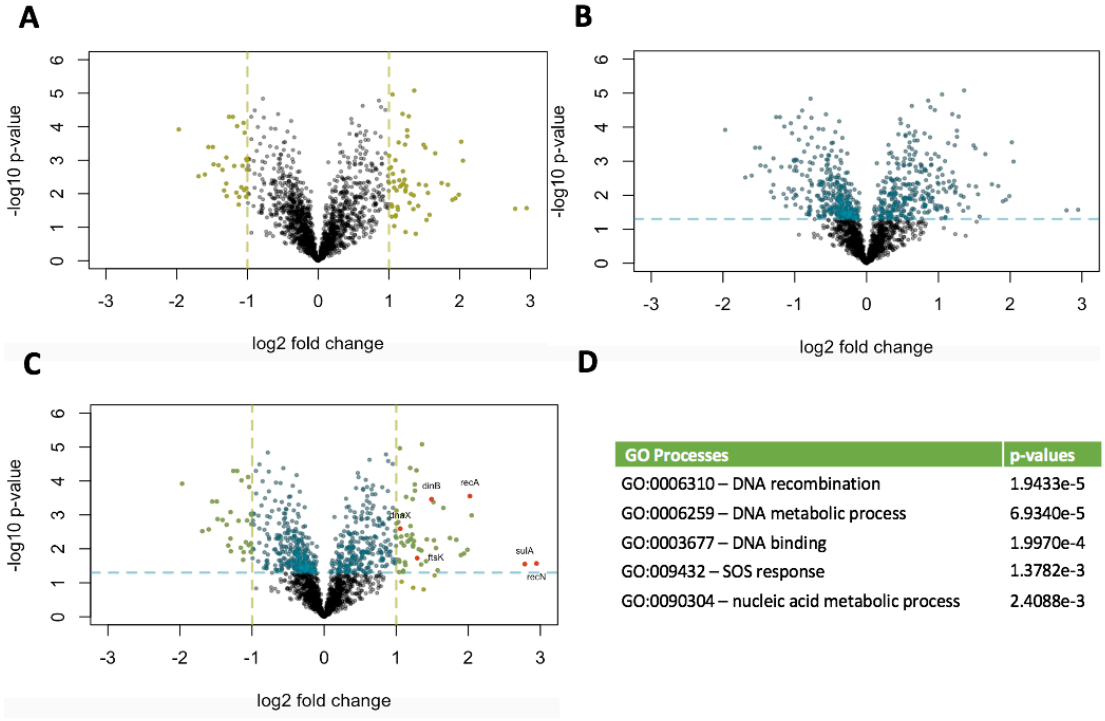
Figure 1. A summary of the spread of differentially regulated promoters in response to antibiotics from various MOA classes

A) The bars depict the average fold change of every promoter in the library in response to chemicals belonging to that MOA class. The horizontal dashed line represents a fold change of 1. B) The y-scale was set to a maximum of 2 to show the spread of downregulated promoters. C) This table summarizes the number of promoters that are differentially regulated for each MOA class, categorized based on magnitude.

3.3 Ciprofloxacin induces the SOS response through DNA damage

Here, we demonstrate a proof of concept example of our investigations into the transcriptional response profiles of the screened chemicals. Ciprofloxacin is a fluoroquinolone antibiotic that induces the SOS response by inhibiting DNA gyrase[14]. To determine differentially regulated promoters post treatment with ¼ MIC of ciprofloxacin, fold change in fluorescence intensity was calculated by dividing the fluorescence intensity values of the drug-treated condition by the untreated control. Welch's unequal variance t-test was performed to determine the statistical significance of the differential phenotypes. Figure 2 depicts the $\log_2$ fold-change in fluorescence intensity in the presence of 1/4 MIC of ciprofloxacin. Promoters were deemed to be differentially expressed in the screen if their fold change value was greater than 2 or less than 0.5 (Figure 2A) and has a p-value of less than 0.05 (Figure 2B).

Indeed, many key SOS response genes, such as *sulA, dinB, recA, recN, ftsK, dnaX,* were upregulated by at least 2-fold relative to the untreated control (Figure 2C). This result showed an enrichment of genes controlled by the LexA transcriptional regulator. The EcoCyc database[36] was used to test for enrichment of gene ontology (GO) terms in biological processes and transcriptional regulation. The results indicate that the ciprofloxacin treatment lead to the upregulation of genes involved in DNA recombination, DNA metabolic processes, DNA binding, the SOS response, and nucleic acid metabolic processes. (Figure 2D).



GO Processes	p-values
GO:0006310 – DNA recombination	1.9433e-5
GO:0006259 – DNA metabolic process	6.9340e-5
GO:0003677 – DNA binding	1.9970e-4
GO:009432 – SOS response	1.3782e-3
GO:0090304 – nucleic acid metabolic process	2.4088e-3

Figure 2. Ciprofloxacin-induced transcriptional response profile. Fold-change in fluorescence intensity in the presence of a sub-inhibitory concentration (1/4 MIC) for ciprofloxacin is shown. Each data point represents the average of 3 biological replicates at their endpoint values. A) The right and left vertical dashed lines mark the cutoff for a fold change of greater than 2-fold or less than 0.5-fold, respectively. Points coloured in yellow show a fold change in gene expression that is either greater than 2-fold or less than 0.5-fold in comparison to the untreated control. B) The horizontal dashed line marks the cutoff for a p-value of 0.05 based on the Welch's t-test. The blue coloured points represent fold change values that are statistically significant (p-value < 0.05). C) Points coloured in red are well characterized SOS responses genes known to be upregulated in response to DNA damage. D) EcoCyc was used to test for GO term enrichment in biological processes and transcriptional regulation.

3.4 Transcription of *rpoS* leads to the upregulation of *dinB*, encoding DNA polymerase IV, in the late exponential phase

As the screen was conducted kinetically, this dataset provides the opportunity to study the induction order of genes, such as *rpoS,* which becomes expressed in the late exponential phase[42]. Figure 3A shows the colony growth overtime for the *E. coli* MG1655 wildtype strain on solid MOPS minimal medium supplemented with 0.4% glucose. The red dashed line represents the time at which the colony reached mid-exponential phase after 3.5 hours. The blue dashed line shows the time at which the colony has passed the exponential phase and transitioned to stationary phase, at 24 hours.

As an example, the transcription of *dinB*, the error-prone DNA polymerase IV, is activated in response to ciprofloxacin induced DNA damage by the sigma factor RpoS in late exponential phase[43]. On the other hand, *recN*, a DNA repair protein, is not under the control of RpoS and is expected to be induced during early log phase[20]. Here, we observe that *recN* is induced by more than 2-fold in the mid-exponential phase (Figure 3B), but *dinB* and *rpoS* show no difference in transcriptional activation during this growth phase. However, as the cells reach stationary phase (Figure 3C), *dinB* becomes induced by 3-fold relative to the untreated control, and *rpoS* is upregulated by 2-fold.

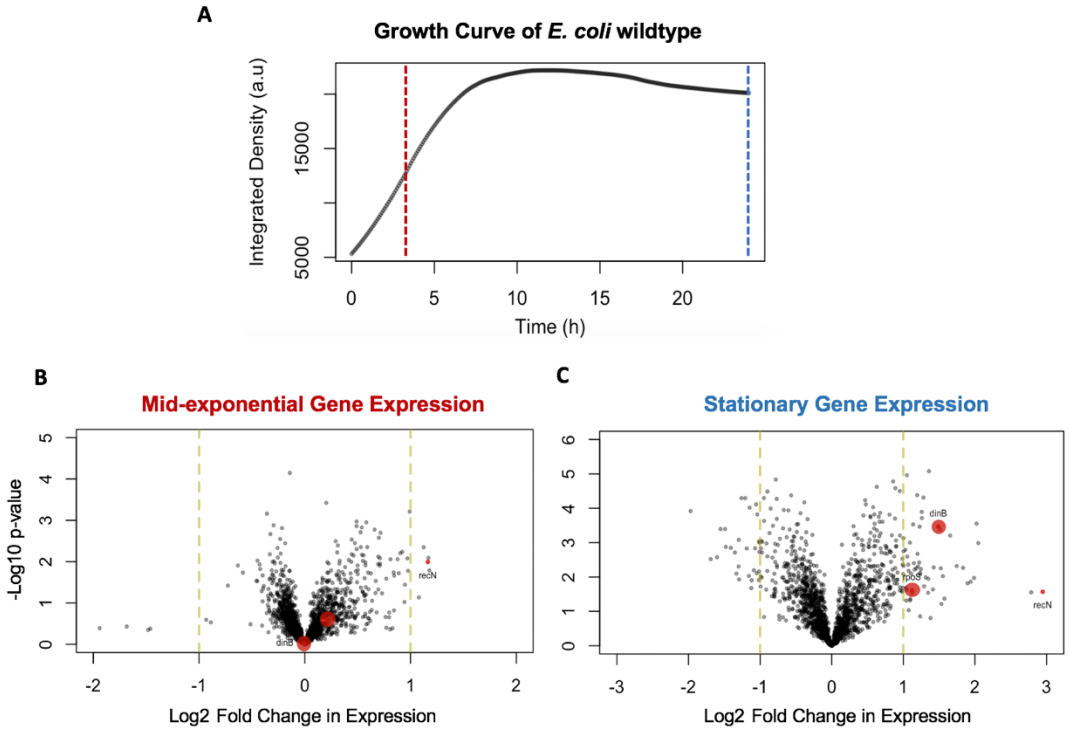


Figure 3. Transcription of *rpoS* leads to the upregulation of *dinB*, encoding DNA polymerase IV, in the late exponential phase. A) The black curve shows the colony growth overtime for an *E. coli* MG1655 wildtype strain on solid MOPS minimal medium supplemented with 0.4% glucose. The red dashed line represents the time at which growth reaches mid-exponential phase after 3.5 hours. The blue dashed line shows the time at which the cells are in the stationary phase, at 24 hours. Fold-change in fluorescence intensity in the presence of a sub-inhibitory concentration (1/4 MIC) for ciprofloxacin is shown at the B) 3.5 h mid-exponential timepoint and C) at the 24 h stationary time point. Each data point represents the average of 3 biological replicates at their endpoint values. The right and left vertical dashed lines in each figure mark the cutoff for a fold change of greater than 2-fold or less than 0.5-fold, respectively. Points coloured in red are the *dinB*, *rpoS*, and *recN* reporter strains.

3.5 Linear discriminant analysis illustrates variations in differentially regulated promoters between classes of compounds

A linear discriminant analysis (LDA) was performed on the endpoint data to examine the variations in transcriptional signatures between the known classes of compound in the training set (Figure 4A). The MOA classes were clearly separated based on transcriptional signals from promoter-reporter strains, including the nutrient biosynthesis inhibitors; this is a class of compounds with great therapeutic potential and is conditionally antimicrobial in nutrient-limited conditions. Each of these classes were separated by variations in gene expression fingerprints based on colony fluorescence. The first three discriminant dimensions explained ~75% (44.5%, 18.5%, and 12.7% respectively) of the variances between the classes (Figure 4B). Particularly, promoters for *dhaM*, *cspA,* and *ygbA* were most important in explaining variations between membrane depolarizing drugs and bacterial cell wall active drugs in the first discriminant dimension LDA1. Promoters for *yeiE*, *kdsB*, and *fhuC* contributed to the separation of drugs targeting folate metabolism, and protein translation inhibitors in the second discriminant dimension LDA2. The third dimension, LDA3, uniquely distanced the esoteric nutrient biosynthesis inhibitors away from the more canonical bioactive drugs. The promoters for *kdsB*, *proS*, *hscC*, and *ycbW* contributed the most to this separation.

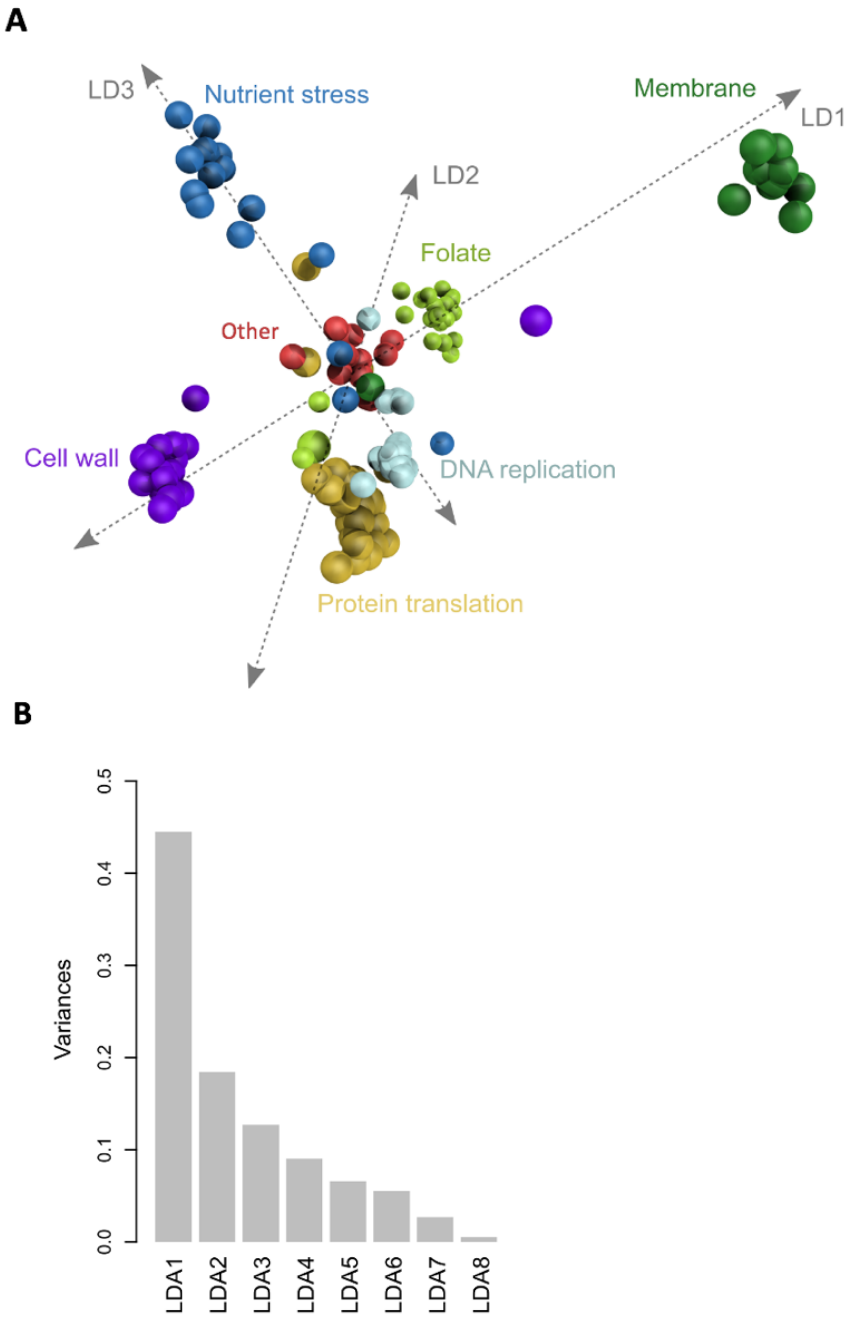


Figure 4. Linear discriminant analysis illustrates variations in differentially regulated promoters between classes of compounds. A) Each MOA class is represented by its own colour and occupies a unique place in this three-dimensional component space. Refer to Table 1 for antibiotics under each classification. B) The first three dimensions of the linear discriminant analysis explains ~75% (44.5%, 18.5%, and 12.7% respectively) of the variances between the classes.

3.6 Highly and uniquely transcribed promoters may act as diagnostic reporters for chemical screens

One area of interest while exploring transcriptional responses to antibiotics is identifying highly and uniquely transcribed promoters that may be used as diagnostic reports in small molecule screens. To this end, promoters were first prioritized using the criteria that they must exhibit a fold change in fluorescence intensity of more than 4-fold in at least one treatment condition relative to the untreated control. A total of 65 promoters were found to satisfy these criteria and were visualized in a heatmap where bright red denotes a fold change of 8-fold or more and black represents no change in fluorescence intensity (Figure 5). Hierarchical clustering was performed for both the list of active promoters and the treatment conditions. At a high level, we observed trends in genes associated with the SOS stress response (*recA, sulA, recN*) in DNA damaging agents, response to oxidative stress (*soxS, fpr*) in superoxide producing compounds, as well as ribosome binding and assembly genes (e.g., *rmf, rsd, rpsJ, rrnB, rpsJ*). Generally, promoters that were induced to a high magnitude in one or a couple of conditions were specifically regulated in those conditions only and were not impacted by other chemicals. The exception to this observation was the ribosome binding and assembly genes, which were induced in many conditions and not specific to certain chemicals or MOA class.

The *sulA* promoter is an example of a reporter that exhibits a high magnitude of induction in response to a specific group of antibiotics (Figure 6). Here, *sulA* was differentially upregulated by all DNA damaging molecules with the exception of novobiocin. It was also upregulated by the folate biosynthesis inhibitor trimethoprim.

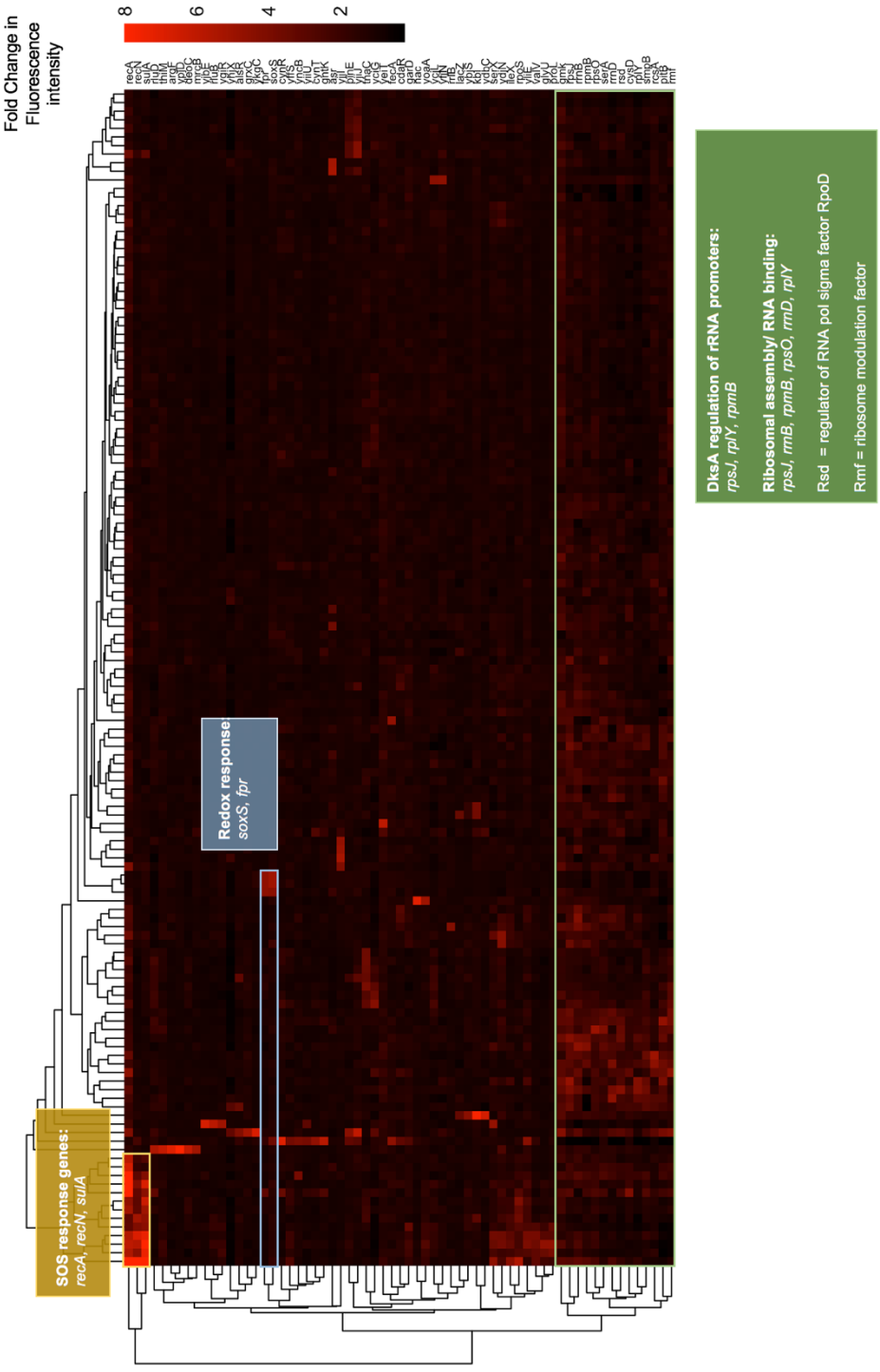


Figure 5. Collection of upregulated promoters with a large magnitude of induction. This collection of the most active promoters was prioritized based on the condition that they exhibit greater than 4-fold increase in fluorescence intensity in at least one treatment condition. All drug treated conditions were normalized to an untreated control. Red represents a fold change of greater than or equal to 8-fold. Black represents no difference in fluorescence intensity between the treated and untreated samples. Hierarchical clustering was performed using the fold change values for both the list of active promoters and treatment conditions.

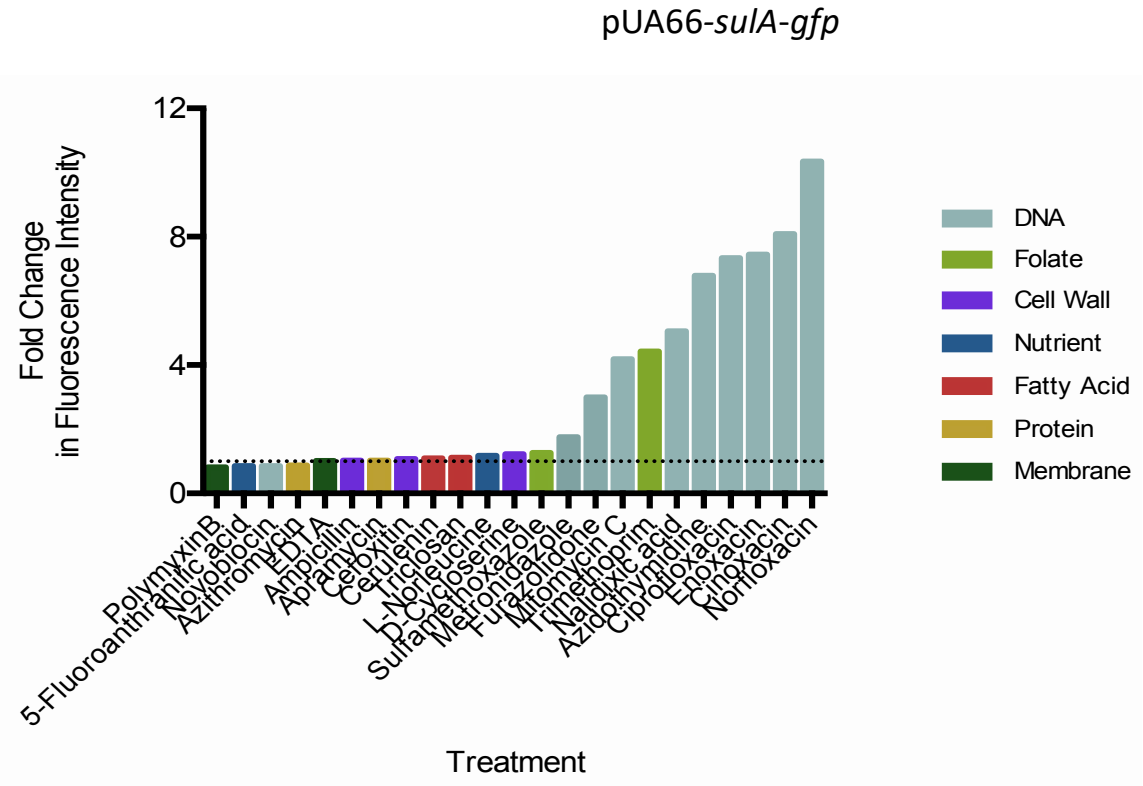


Figure 6. The *sulA* promoter is uniquely induced to a high magnitude by chemicals that primarily perturb DNA synthesis. Fold change in fluorescence intensity for the *sulA* reporter strain against a panel of drug treatment conditions. Drug treatments are colored based on MOA class and listed in ascending order. If an antibiotic was screened at multiple concentrations, the concentration resulting in the largest fold change, either increase or decrease in fluorescence, was used for this analysis. Antibiotics belonging to other MOA classes are shown here as representative examples of that class. The dashed horizontal line indicates a fold change of 1.

Representatives chemicals from other MOA classes show no changes in fluorescence intensity for the *sulA* promoter.

The *recA* gene is also induced in response to DNA damaging agents. It is one of the first genes to be activated in the DNA repair response and can be activated by weak inducers of DNA damage[20]. Here, we observe that the *recA* promoter is generally induced by antibiotics from all MOA classes to a certain extent (Figure 7). However, the antibiotics that induces *recA* to the highest magnitudes are chemicals that perturb DNA synthesis. The cell wall biosynthesis inhibitors ampicillin and D-cycloserine also induce *recA* to a moderate extent, indicating DNA damage as a secondary effect.

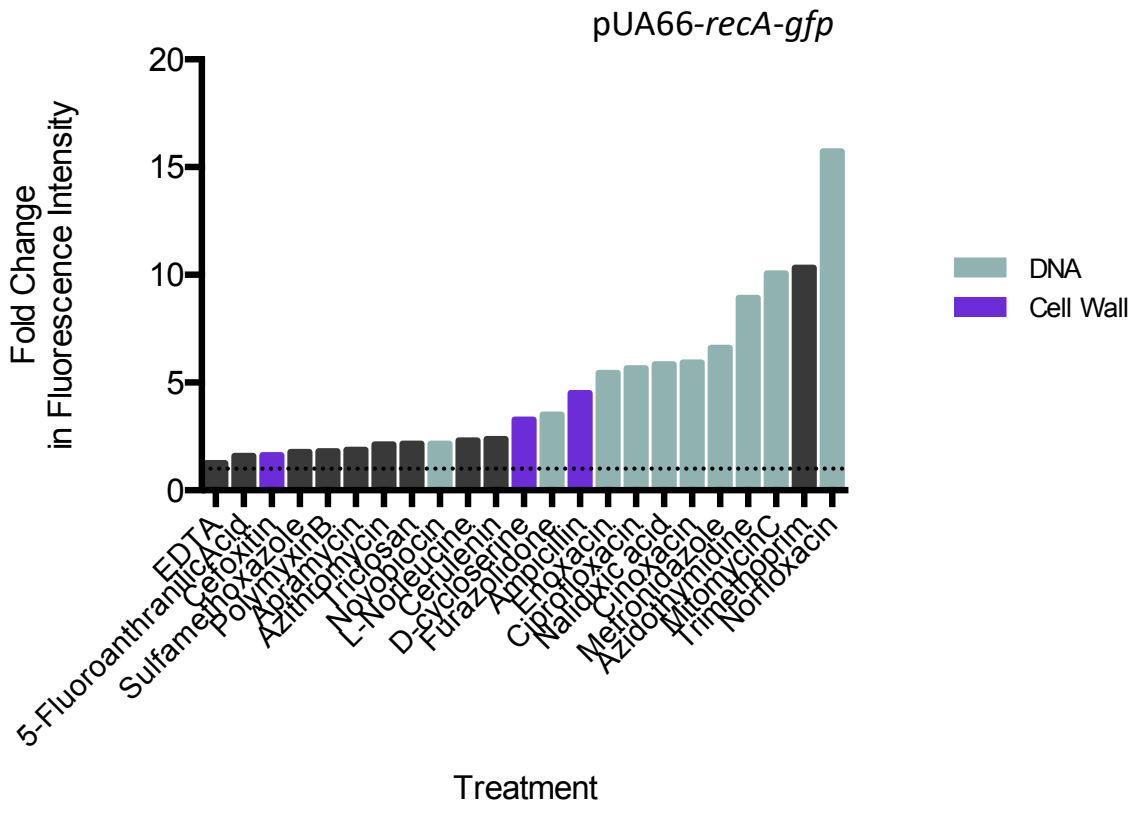


Figure 7. Defective cell wall biosynthesis caused by ampicillin and D-cycloserine induces error-free DNA repair via induction of *recA*. Fold change in fluorescence intensity for the *recA* reporter strain against a panel of drug treatment conditions. Drug treatments are colored based on MOA class and listed in ascending order. Black represents compounds from other MOA classes. If an antibiotic was screened at multiple concentrations, the concentration resulting in the largest fold change, either increase or decrease in fluorescence, was used for this analysis. Antibiotics belonging to other MOA classes were shown here as representative examples of that class. The dashed horizontal line indicates a fold change of 1.

The *fpr* gene encodes the flavodoxin/ferredoxin-NADP+ reductase, which is induced in response to oxidative stress[44]. Here, we observe that the *fpr* promoter is induced to a high magnitude by paraquat, trimethoprim, and A22, an inhibitor of MreB (Figure 8). Paraquat[45] and trimethoprim[46] are both known to be superoxide producers. Chemicals from other MOA classes did not induce this promoter; the average fold change without these three chemicals is 1.01.

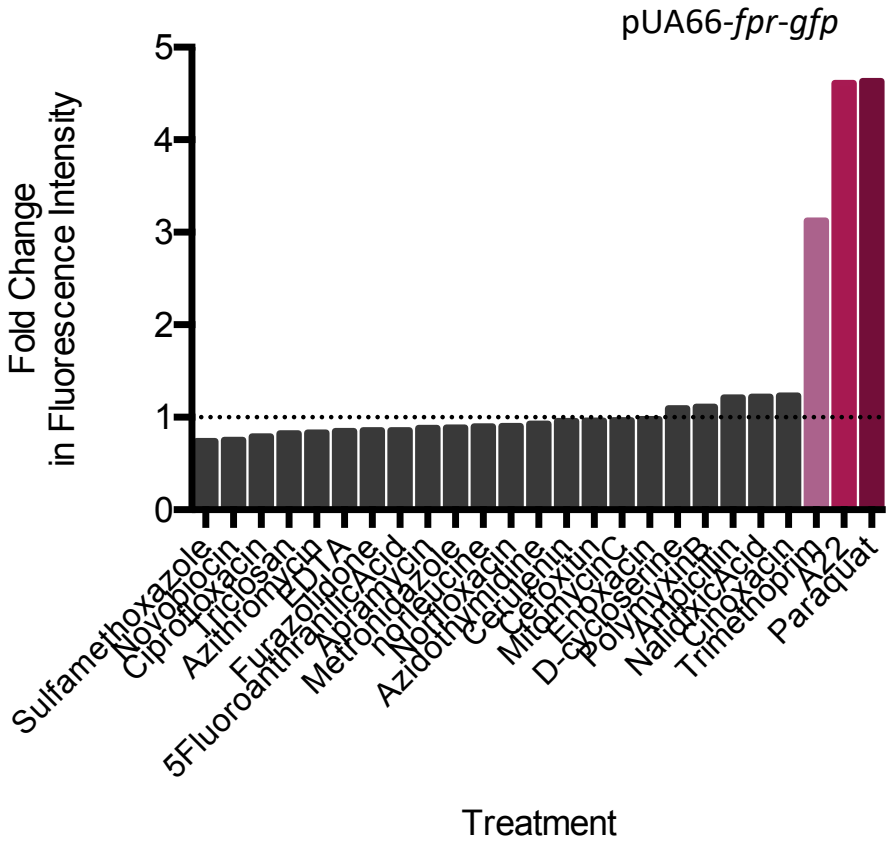


Figure 8. Transcription of the *fpr* promoter is induced in response to superoxide production. Fold change in fluorescence intensity for the *fpr* reporter strain against a panel of drug treatment conditions. If an antibiotic was screened at multiple concentrations, the concentration resulting in the largest fold change, either increase or decrease in fluorescence, was used for this analysis. Antibiotics that resulted in a large fold change in fluorescence intensity are coloured in pink. The average fold change for all drug treatments excluding paraquat, A22, and trimethoprim is 1.01. The dashed horizontal line indicates a fold change of 1.

Lastly, RcsA is a positive transcriptional regulator of capsular polysaccharide synthesis[47]. It has been shown to be induced by certain cell wall biosynthesis inhibitors[10] and outer membrane perturbants[48] in rich media. Here, we show that the penam group of cell wall biosynthesis inhibitors (mecillinam, penicillin G, ampicillin, and D-cycloserine) induced the *rcsA* promoter to the highest magnitude. The cephem group of antibiotics (cefazolin, cefadroxil, cefoxime, cefmetazole, cefoxitin) did not show conserved induction of *rcsA*. Additionally, we did not observe *rcsA* induction with membrane perturbants as hypothesized.

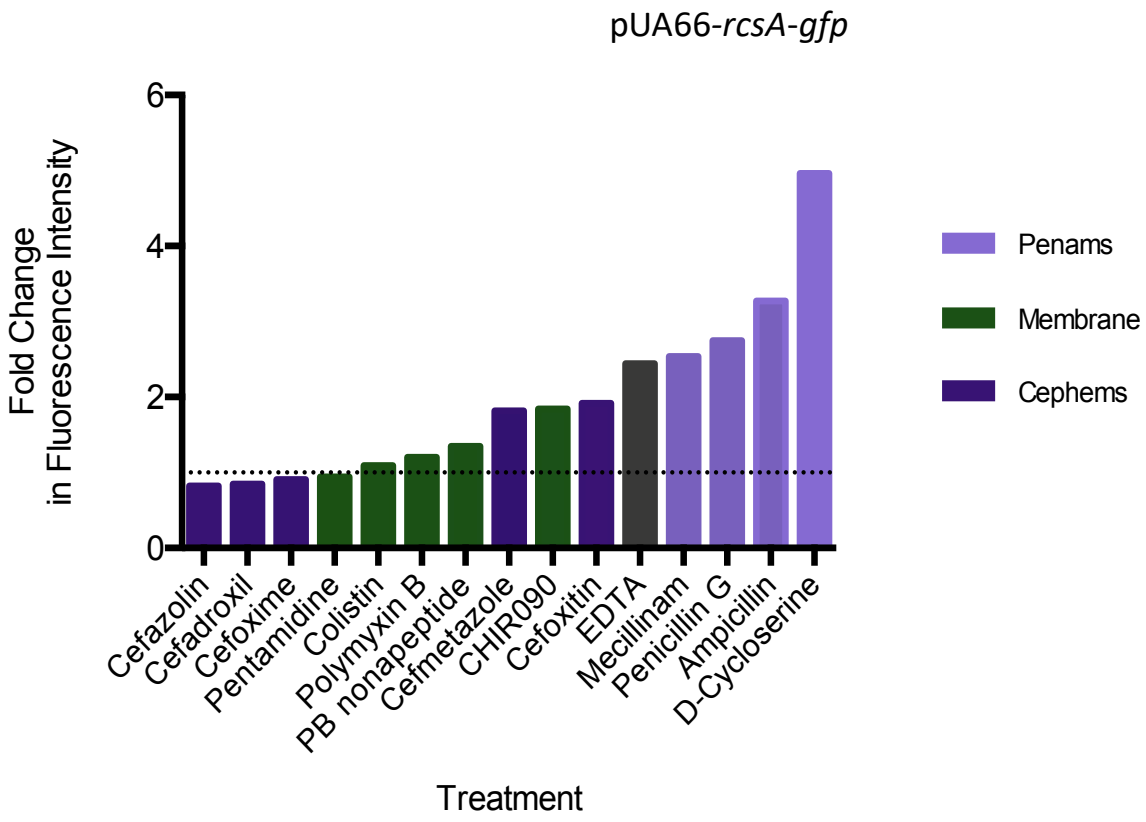


Figure 9. Transcription of the *rcsA* promoter is induced to the highest magnitude by penicillin derivatives. Fold change in fluorescence intensity for the *rcsA* reporter strain against a panel of drug treatment conditions. Drug treatments are colored based on MOA class and listed in ascending order. If an antibiotic was screened at multiple concentrations, the concentration resulting in the largest fold change, either increase or decrease in fluorescence, was used for this analysis. The dashed horizontal line indicates a fold change of 1.

3.7 A deep learning model for predicting antibiotic mechanism of action

A deep learning network was built using the Keras package in R, with Tensorflow as the backend (Figure 10A). A 10-layer model was constructed consisting of: 2D convolution (64 filter, 5x5 kernel, rectified linear unit (relu) activation), 0.25 dropout, 2D convolution (128 filter, 3x3 kernel, relu activation), 0.25 dropout, 2D pooling (pool size 4), 0.25 dropout, flattened layer, densely connected layer (50 unit, relu activation), 0.25 dropout, densely connected layer (softmax activation). The input layer had 6,144 neurons, the number of colonies in a 64x96 array. The network was compiled with the Adam optimizer with a binary cross-entropy loss function. Accuracy was visualized alongside loss while the model was compiled, and both measures levelled out after 10 epochs. Internal validations were done with an 80/20 random split of the data. 80% of the data acted as the training set in which the model learns from and the remaining 20% of the data acts as the validation set used to monitor the accuracy. Network accuracy was about 98% (Figure 10B) and loss approaching 0.25% (Figure 10C).

Predictions using this system were dependent on patterns of transcriptional response to each chemical in each agar plate. The accuracy of the model was assessed by first generating new data for drugs that were either already present or absent in the training set, then computing a MOA prediction for these compounds using the model (Figure 11). Cefmetazole and polymyxin B were used as “known” test molecules to determine the accuracy of the model. These compounds had already been included in the training set. The percent probability score for these two chemicals were 95% and 100%, respectively. The fluroquinolone cinoxacin was used as an ‘unknown’ test molecule to

identify what it would match with in the existing training data. Cinoxacin was not included in the training set and was assayed completely independently of the training set. Indeed, the best prediction for the 'unknown' was a sub-inhibitory concentration of enoxacin (~80% prediction confidence), a structurally similar fluoroquinolone.

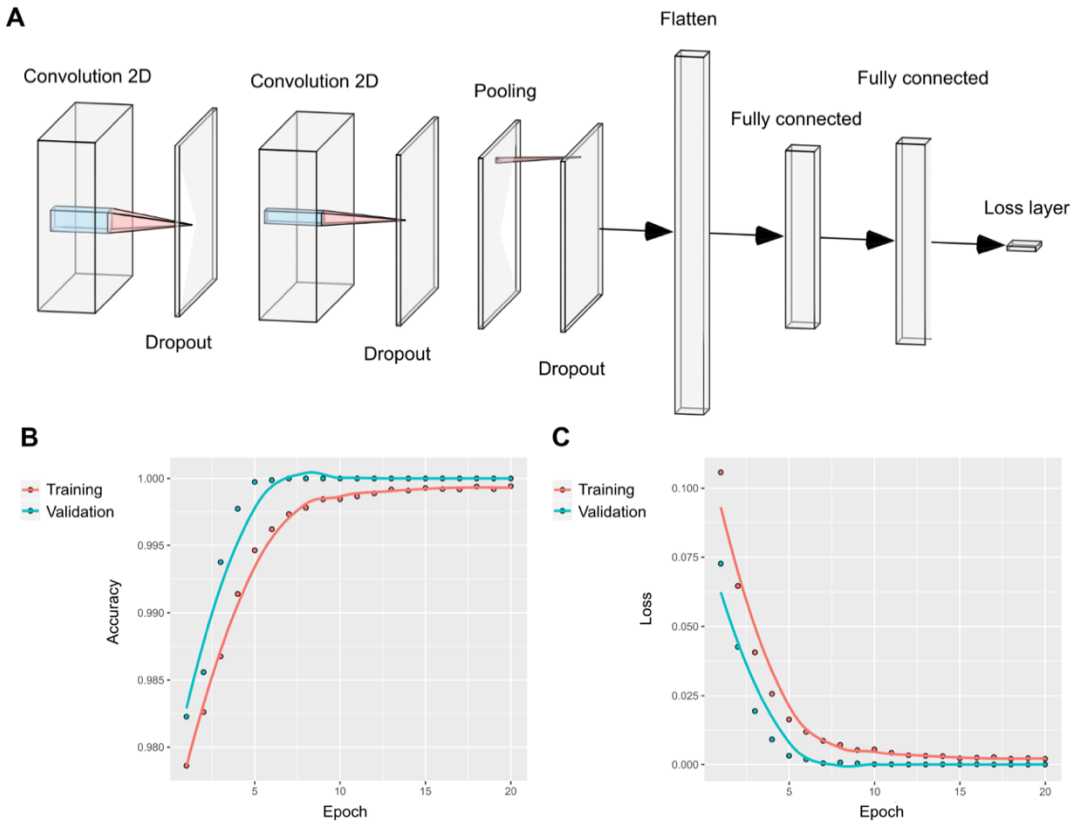
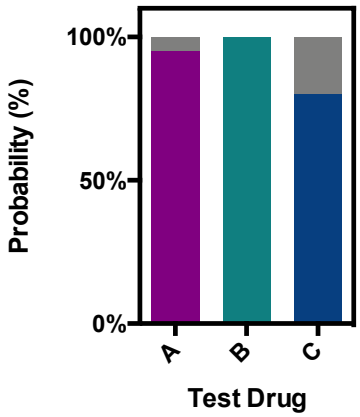


Figure 10. A neural convolutional network for predicting the mechanism of action of unknown chemicals based on a training set of known transcriptional profiles. A) A deep learning network was built using the Keras package in R, with Tensorflow as the backend. A 10-layer model was constructed consisting of: 2D convolution, 0.25 dropout, 2D convolution, 0.25 dropout, 2D pooling, 0.25 dropout, flattened layer, densely connected layer, 0.25 dropout, densely connected layer (softmax activation). The network was compiled with the Adam optimizer with a binary cross-entropy loss function. B) Accuracy was visualized alongside C) loss, and both measures levelled out after 10 epochs. Internal validations were done with an 80/20 split of the data.



Chemical	Prediction	Identity
A	95% Cefmetazole, 5% Gentamicin	Cefmetazole
B	100% Polymyxin B	Polymyxin B
C	80% Cinoxacin, 20% Azithromycin	Enoxacin

Figure 11. Convolutional neural network predictor model can recognize and identify certain test compounds. A single-blind study to test the model for its ability to recognize and identify unknown test drugs. The model correctly predicted the identity of cefmetazole (A) with a 95% probability (purple) and polymyxin B (B) with 100% probability (teal). The model predicted the identity of enoxacin (C) with 80% probability (blue) treated sample, which was not included in the training set, to its close derivative, cinoxacin.

3.8 An antimetabolite molecule MAC168425 with an unknown MOA was predicted to have a MOA similar to that of trimethoprim

In a small molecule screen by Zlitni et al[49], a novel compound, MAC168425 (Figure 12A), was discovered to exhibit inhibitory activity against wildtype *E. coli* in a nutrient limited medium. This molecule had an MIC of 8 μg/mL against wildtype *E. coli* in MOPS minimal medium with 0.4% glucose as the carbon source (Figure 12B). The inhibitory activity of MAC168425 was suppressed by supplementing glycine into the growth medium (Figure 12B). The MIC of MAC168425 increased by 2- fold when 50-200 μg/mL of glycine was supplemented into the medium. When 400 μg/mL of glycine was supplemented, the MIC reached 32 μg/mL, resulting in a 4-fold suppression. This result suggests that MAC168425 may be targeting the glycine pathway.

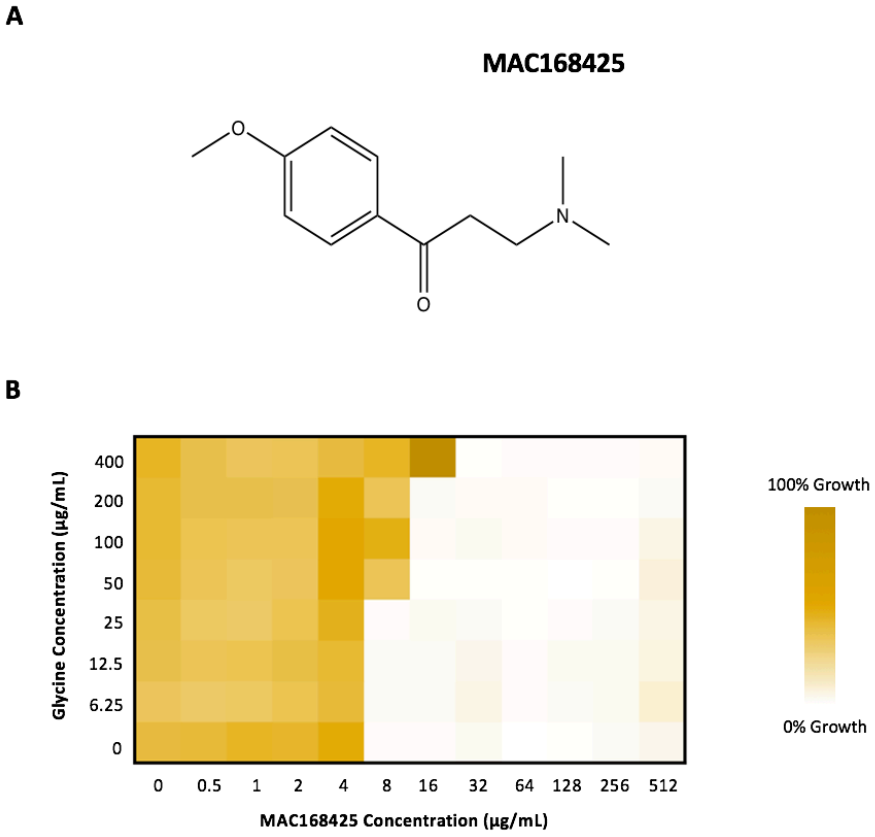


Figure 12. MAC168425 activity is suppressed by supplementation of glycine
A) The chemical structure of MAC168425. B) Checkerboard analysis of the antagonist relationship between glycine and MAC168425. Dark yellow indicates maximal growth in this experiment and white denotes no bacterial growth. The columns contain increasing concentrations of MAC168425. The rows contain increasing concentrations of glycine supplemented into the MOPS medium.

To further test the deep learning model built in this study, we generated a prediction probability output for MAC168425 (Figure 13). The model's prediction output for MAC168425 is as follows: 80% trimethoprim, 10% 6-mercaptopurine, 5% nalidixic acid, 5% 6-diazo-5-oxo-norleucine. This suggests that MAC168425 may have a MOA that is the same or closely related to that trimethoprim. This became our new hypothesis for MAC168425 as we continued to explore its mechanism of action.

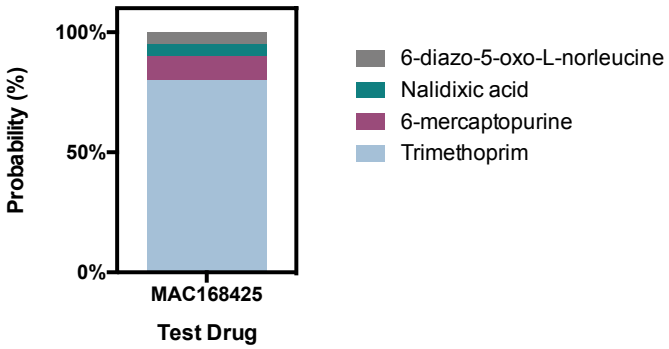


Figure 13. Convolutional neural network predictor model predicts that MAC168425 has a mechanism of action closely related to trimethoprim. A single-blind study to compute a MOA prediction for a compound of unknown MOA, MAC168425. The model prediction in % probability is as follows: 80% trimethoprim, 10% 6-mercaptopurine, 5% nalidixic acid, 5% 6-diazo-5-oxo-L-norleucine.

3.9 Overexpression of FolA does not lead to resistance against MAC168425

One approach for drug target identification is based on the hypothesis that target overexpression can confer resistance to the antibiotic. When higher copies of the protein target are present, higher concentrations of the antibiotic will be required to reduce target activity and cause growth inhibition or cell death[50]. Previously, overexpression of FolA has been shown to result in resistance to trimethoprim[50]. To determine whether MAC168425 also inhibits the same protein target as trimethoprim, the enzymatic target of trimethoprim, FolA, was overexpressed intracellularly using the pCA24N-*folA* plasmid from the ASKA collection[38]. Indeed, we observed that overexpression of FolA led to resistance to trimethoprim; the MIC of trimethoprim increased by at least 32-fold in the overexpression strain compared to the empty vector strain (Figure 14A). However, overexpression of FolA did not confer resistance to MAC168425, as no difference in the MIC between the overexpression and empty vector strain was observed (Figure 14B).

This result suggests that FolA was likely not the enzyme target of MAC168425. One possible explanation for this is that there is currently no probe in the dataset that target the same enzyme as MAC168425. Thus, the model may be suggesting trimethoprim as the chemical that resembles MAC168425 the most in terms of transcriptional response. Nonetheless, we hypothesize that the MOA of MAC168425 could still be closely related to the folate biosynthesis or folate utilization pathway.

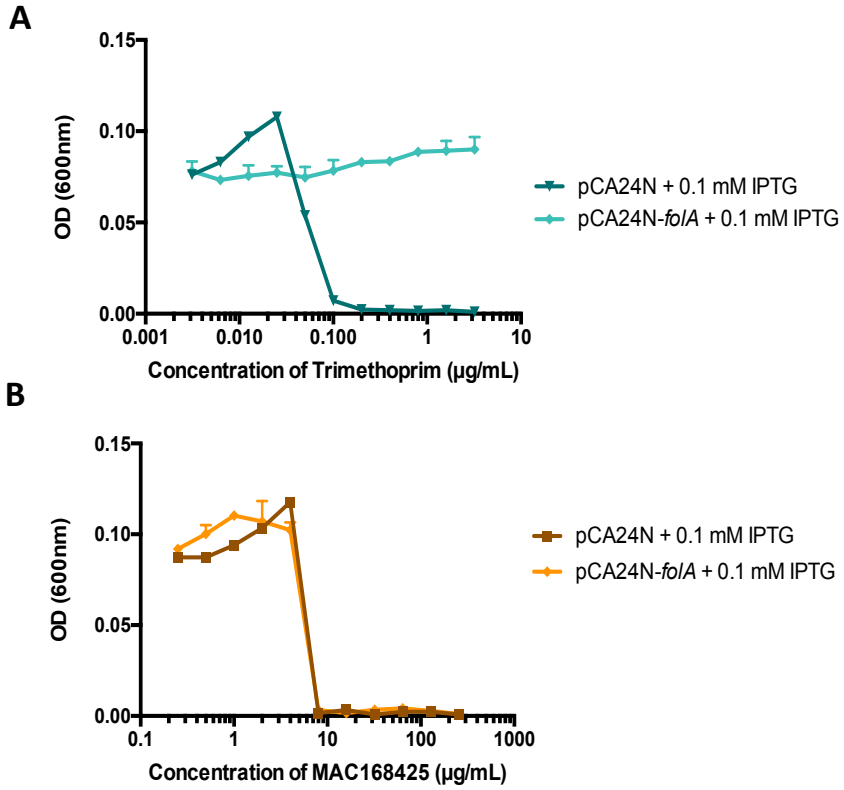


Figure 14. Overexpression of FolA suppresses the activity of trimethoprim but not MAC168425. Induction of FolA overexpression using 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) was performed for pCA24N (empty vector control) and pCA24N-*folA*. A) pCA24N (dark teal) and pCA24N-*folA* (light teal) were grown in the presence of 2-fold dilutions of trimethoprim. B) pCA24N (light orange) and pCA24N-*folA* (dark orange) were grown in the presence of 2-fold dilutions of MAC168425. Experiments were performed on three biological replicates. The data shown represents the average of 3 replicates ± standard deviation.

3.10 MAC168425 activity is synergistic with sulfamethoxazole and antagonist against trimethoprim

Checkerboard assays were performed with sulfamethoxazole and trimethoprim to further probe the relevance of folate biosynthesis and utilization for the activity of MAC168425. Figure 15A shows a known synergistic interaction between sulfamethoxazole and trimethoprim[51]; this is a positive control for the experiment. When sulfamethoxazole and MAC168425 were assayed in combination (Figure 15B), there was also a synergistic interaction in terms of growth inhibition. When trimethoprim and MAC168425 were tested in combination, there was an antagonist interaction in terms of growth inhibition (Figure 15C). These interactions suggest that MAC168425 could be 1) also targeting DHFR as is trimethoprim but using a different mechanism, 2) targeting dihydropteroate synthase (FolP), the target of sulfamethoxazole at a different site, or 3) inhibiting an earlier or later step in a common metabolic pathway.

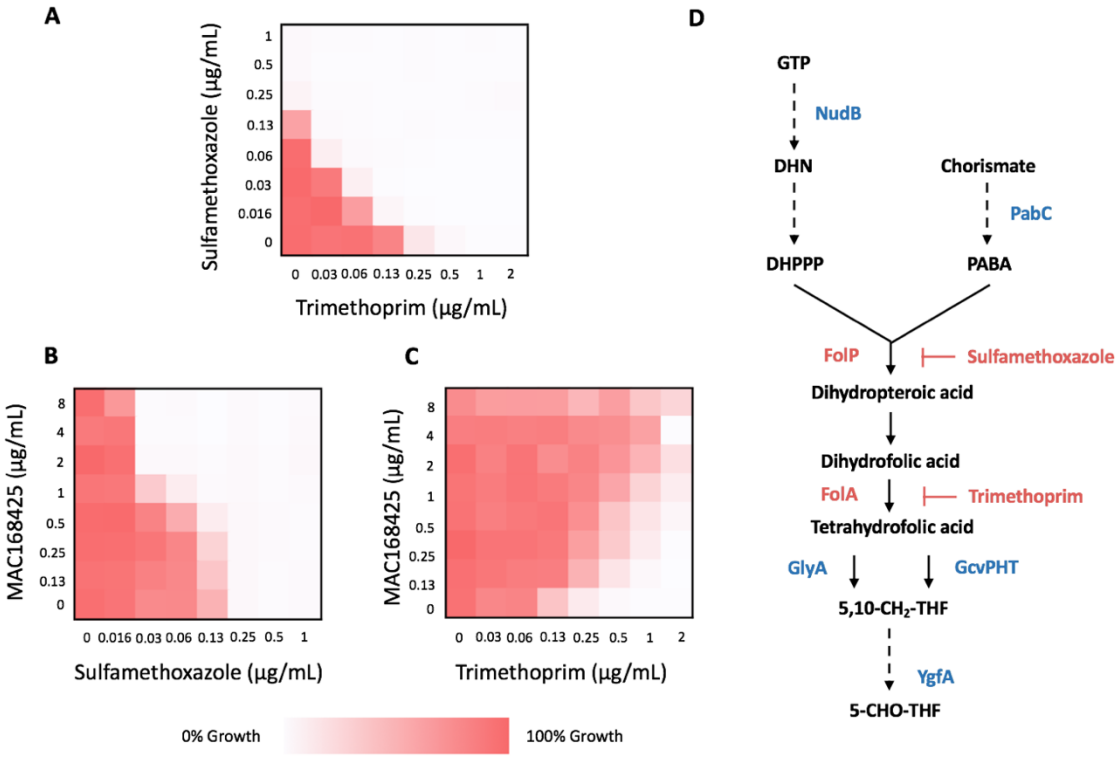


Figure 15. MAC168425 activity is synergistic with sulfamethoxazole and antagonist against trimethoprim. All checkerboard assays were performed with *E. coli* BW25113 wildtype. Red indicates maximal growth in this experiment and white denotes no bacterial growth. The columns contain increasing concentrations of trimethoprim or sulfamethoxazole. The rows contain increasing concentrations of MAC168425 of sulfamethoxazole. A) The synergistic interaction of sulfamethoxazole and trimethoprim. B) The synergistic interaction of MAC168425 and sulfamethoxazole. C) The antagonistic interaction between MAC168425 and trimethoprim. D) The steps of folate biosynthesis involving the inhibition of sulfamethoxazole and trimethoprim. Dashed arrows represent more than one biosynthetic step. DHN: 7,8-dihydroneopterin; DHPPP: 7-8-dihydropterin pyrophosphate, PABA: *para*-aminobenzoic acid, THF: tetrahydrofolate.

3.11 Deletion mutants involved in folate biosynthesis and glycine transport exhibit no change in sensitivity to MAC168425

To determine whether the synergistic and antagonistic phenotypes of MAC168425 can be explained by its involvement in a related step of folate biosynthesis (Figure 15D), a selection of single deletion mutants were tested for their sensitivity to MAC168425. Mutations in genes involved in folate biosynthesis, such as *ΔnudB, ΔgcvP, ΔygfA,* showed no change in sensitivity to MAC168425, with the exception of *ΔgcvT* and *Δlrp* being 2-fold more sensitive to MAC168425.

While investigating mutant sensitivity to MAC168425, we also explored the hypothesis that excess amount of glycine can suppress the activity of MAC168425 if they were competing for the same transporters for entry into the cell. To this end, deletion mutants in *gcvB, cycA,* and *sstT* were tested for changes in sensitivity (Table 4). However, there was no change in MIC for these strains, indicating that competition for entry to the cell was likely not the reason behind the glycine suppression phenotype.

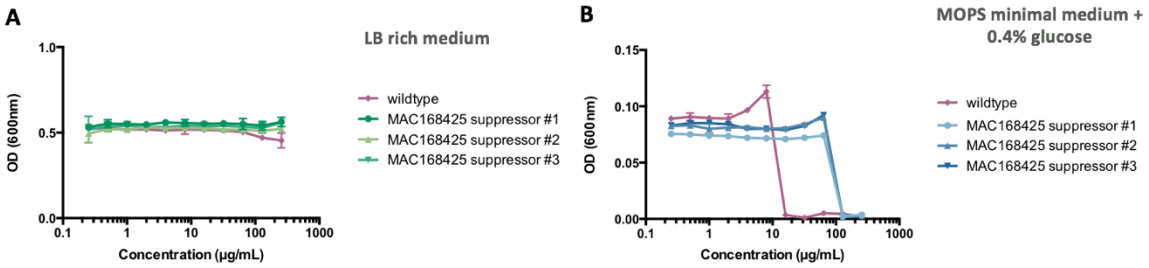
Table 4. Minimum inhibitory concentrations of MAC168425 for all strains tested in this study. All single gene deletion strains are from the Keio collection. MICs were determined in MOPS medium supplemented with 0.4% glucose.

Strain	Protein Description	MIC (μg/mL)	Fold Change
wildtype		16	-
Δ*csrD::kan*	Carbon storage regulator	16	1
Δ*cycA::kan*	Serine/alanine/glycine/:H^+ symporter	16	1
Δ*gcvA::kan*	Repressor of glycine cleavage system	16	1
Δ*gcvB::kan*	sRNA regulator	16	1
Δ*gcvH::kan*	Glycine cleavage system H-protein	16	1
Δ*gcvP::kan*	Glycine cleavage system P-protein	16	1
Δ*gcvT::kan*	Glycine cleavage system T-protein	**32**	**2**
Δ*lrp::kan*	Leucine-responsive regulatory protein	**8**	**2**
Δ*nudB::kan*	Dihydroneopterin triphosphate diphosphatase	16	1
Δ*pabC::kan*	Aminodeoxychorismate lyase	**1**	**1**
Δ*sstT::kan*	Serine/threonine transporter	16	1
Δ*ygfA::kan*	5-CHO-THF cyclo-ligase	16	1
Δ*ygfZ::kan*	THF-dependent protein	**4**	**4**

3.12 MAC168425 resistant mutants shows 8-fold resistance to MAC168425 and no cross resistance to other antibiotics.

To further investigate the target of MAC168425 through target modification resistance mechanisms, suppressor mutants were generated on MOPS minimal agar plates supplemented with 4x the MIC of the drug. Three biological replicates were confirmed to be resistant to MAC168425 after passaging on LB medium with no drug. MAC168425 does not have any inhibitory activity in LB rich medium (Figure 16A). In MOPS minimal

+ 0.4% glucose medium (Figure 16B), resistant mutants showed an 8-fold resistance to MAC168425 relative to wildtype. Furthermore, these mutants did not show any cross resistance against a panel of other antibiotics, including trimethoprim and sulfamethoxazole (Figure 16C).



C

Antibiotic	MAC168425 Suppressor #1 (µg/mL)	MAC168425 Suppressor #2 (µg/mL)	MAC168425 Suppressor #3 (µg/mL)	*E. coli* BW25113 wildtype (µg/mL)	Fold Change
Cefmetazole	1	1	1	1	-
Gentamicin	0.03	0.03	0.03	0.03	-
Rifampicin	8	8	8	8	-
Ampicillin	4	4	4	4	-
Ciprofoxacin	0.008	0.008	0.008	0.008	-
Colistin	0.25	0.25	0.25	0.25	-
Erythromycin	256	256	256	256	-
Tetracycline	2	2	2	2	-
Trimethoprim	0.5	0.5	0.5	0.5	-
Sulfamethoxazole	0.25	0.25	0.25	0.25	-
MAC168425	128	128	128	16	**8**

MOPS minimal medium + 0.4% glucose

Figure 16. MAC168425 resistant mutants shows 8-fold resistance to MAC168425 and no cross resistance to other antibiotics. *E. coli* wildtype and the MAC168425 resistant mutants were tested in the presence of increasing concentrations of MAC168425 in A) LB rich medium and B) MOPS minimal + 0.4% glucose medium. C) Drug susceptibility analysis of wildtype and each MAC168425 suppressor mutant background. The minimum inhibitory concentration for each drug is shown in the coloured columns under each strain background.

3.13 MAC168425 resistant mutants show differential expression of *csrD* and *yhdH*

Genomic DNA of these MAC168425 resistant mutants and the parent wildtype strain were isolated for whole genome sequencing. Illumina sequencing libraries were constructed using a Nextera DNA libray prep kit. Paired end sequencing (2x150 bp) at ~150x coverage was carried out on an Illumina NextSeq 550 system. Adapter sequences were trimmed, and mutations were identified using the breseq pipeline[52]. Mapping the MAC168425 resistant strains against the parental wildtype strain revealed the presence of a mobile insertion sequence, IS5, in the intergenic region between *csrD* and *yhdH* at position 3,396,784 in the *E. coli* BW25113 genome (Figure 17A). The insertion element was positioned upstream of both *csrD* and *yhdH*. CsrD is a regulator of the carbon storage system of *E. coli* through RNase E-mediated degradation of the small RNAs, CsrB and CsrC[53]. YhdH is a putative acrylyl-CoA reductase; its physiological role is otherwise poorly characterized in *E. coli*[54].

IS5 is a 1,195 bp mobile bacterial DNA element that can transpose to other sites on the *E. coli* genome and result in genetic arrangements[55]. Previously, IS5 has been shown to regulate transcriptional activity through insertion into the regulatory regions of various genes[565758]. To investigate the effects of IS5 insertion on *csrD* and *yhdH*, RT-qPCR was performed on the parent wildtype and three mutant strains. Indeed, the insertion of IS5 upstream of *csrD* and *yhdH* resulted in differential gene expression. Relative to the wildtype strain, the resistant mutants show ~40% reduction in *csrD* transcript level (Figure 17B) and ~14-fold increase in *yhdH* transcript level (Figure 17C).

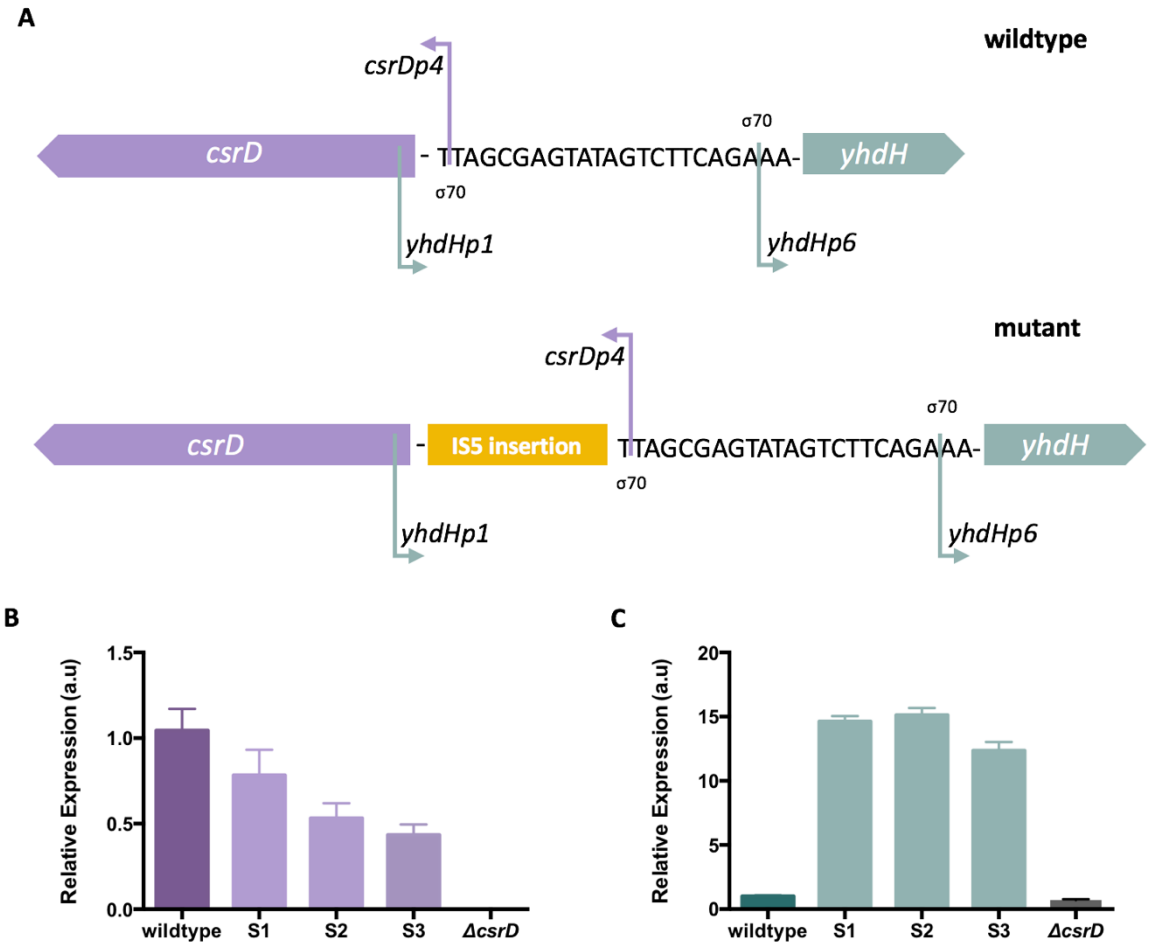


Figure 17. *E. coli* confers resistance to MAC168425 through upregulation of YhdH. A) Mapping the MAC168425 suppressor strains against the parental *E. coli* BW25113 wildtype strain revealed the presence of a mobile insertion element IS5 in the intergenic region between *csrD* and *ydhH*. All three biologically resistant mutants possessed this insertion element. B) RT-qPCR shows ~40% reduction in *csrD* transcript level. S1, S2, and S3 denotes three biological replicates. Error bars represent the error for three technical replicates. C) RT-qPCR shows ~14-fold increase in *yhdH* transcript level. S1, S2, and S3 denotes three biological replicates. Error bars represent the error for three technical replicates.

3.14 *E. coli* confers resistance to MAC168425 through upregulation of YhdH

To determine the effects of differential *csrD* and *yhdH* expression on developing resistance to MAC168425, a *csrD-null* mutant and an YhdH overexpression strain were tested for their sensitivity to MAC168425. A single deletion of *csrD* did not show any changes in sensitivity to MAC168425; the MIC remained the same between the mutant and wildtype strain (Table 4). Conversely, when YhdH was overexpressed intracellularly in a wildtype background, the cells become 8-fold more resistant to the effects of MAC168425; the MIC in the resistant strains was 128 μg/mL (Figure 18). This suppression phenotype resembles the resistance phenotype we observed in the spontaneous resistant mutants.

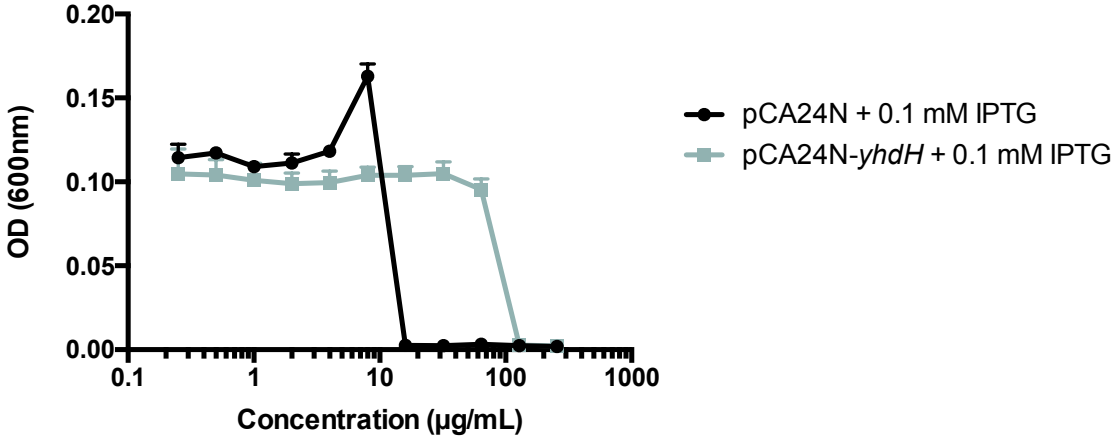


Figure 18. Overexpression of YhdH results in 8-fold suppression of MAC168425 activity
Induction of YhdH overexpression using 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) was performed for pCA24N (empty vector control) and pCA24N-*yhdH*. pCA24N (black) and pCA24N-*yhdH* (grey) were grown in the presence of 2-fold dilutions of MAC168425. The data shown represents the average of 3 biological replicates ± standard deviation.

3.15 MAC168425 is implicated in glycine cleavage and folate-dependent processes

Although overexpression of YhdH confers resistances to MAC168425, it is likely not the enzyme target of MAC168425 because it is not an essential enzyme required for growth in minimal medium. As such, overexpression of YhdH may be a result of compensating for the actual inhibitory activity of MAC168425 or is co-transcribed with the target of MAC168425 by a common regulator.

YhdH is a putative acrylyl-coA reductase and is hypersensitive to acrylate and 3-hydroxyproprionate[54]. However, its physiological role in *E. coli* is otherwise poorly characterized. Homologs of YhdH exist in other bacterial taxa. Namely, the AcuI homolog in the Roseobacter group of marine bacteria degrades acrylate through catabolism of dimethylsulfoniopropionate (DMSP), an abundant nutrient source and anti-stress molecule released by marine plankton[54]. Similar to *E. coli yhdH* mutants, mutants lacking *acuI* are hypersensitive to the inhibitory effects of acrylate[54].

It has been shown that many *acuI*-like genes are located in close proximity with other DMSP catabolism genes such as *dmdA*. Of the 37 searchable Roseobacter strains that contains *acuI,* 26 encode a DmdA demethylase; 24 of these *dmdA* genes are located downstream of *acuI* and are co-transcribed[54]. The acrylate sensitivity of *E. coli ⊿yhdH* can be corrected by cloning *acuI* from bacteria whose *acuI* gene is closely linked to *dmdA*. *E. coli* does not contain the *dmdA* gene or catabolize DMSP, but this close association prompted investigations into *E. coli* protein homologs of DmdA.

DmdA is a member of the aminomethyltransferase-like protein family[59]. In terms of protein homology, it shares structural similarity with the proteins GcvT[60] and YgfZ[61].

GcvT is a member of the glycine cleavage system responsible for the oxidation of glycine[62]. YgfZ has been implicated in the synthesis and repair of iron-sulfur clusters[63], response to oxidative stress[64], and tRNA modification[65]. Interestingly, all these proteins are all folate-binding or folate-dependent enzymes[61,62].

To determine the effects of MAC168425 on GcvT and YgfZ, dose response assays were performed with *gcvT* and *ygfZ*-null mutants. The *gcvT*-null mutant is 2-fold more sensitive to MAC168425 relative to wildtype (Table 4) and the *yhdH*-null mutant is 4-fold more sensitive (Figure 19). This interaction with glycine cleavage and folate-dependent enzymes may explain the glycine suppression phenotype we observed previously, as well as the trimethoprim prediction output from the deep learning model.

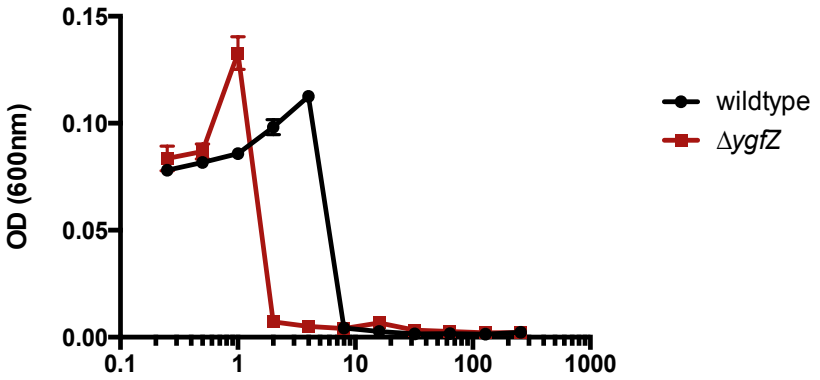


Figure 19. The *ygfZ*-null mutant is 4-fold more sensitive to the activity of MAC168425. Dose response curve of MAC168425 against *E. coli* wildtype (black) and *ΔygfZ* (red). Each strain grown in the presence of 2-fold dilutions of MAC168425. The data shown represent the average of 3 technical replicates ± standard deviation.

CHAPTER 4: DISCUSSION

Here, we investigate the transcriptional response of *E. coli* when challenged with sub-inhibitory concentrations of antibiotics spanning several antibiotic classes. Using a transcriptional promoter-reporter library[31] and the PFIbox[32], we acquired images of fluorescence transcriptional phenotypes in solid media arrays in a high-throughput and inexpensive manner. This provides a comprehensive look at the global promoter activity in *E. coli* by means of promoter-reporter fusion constructs. Although there have been previous investigations aimed at leveraging gene expression profile as a predictor of antibiotic MOA, this is the first study, to our knowledge, to explore this objective in a Gram-negative bacterium, *E. coli,* under nutrient limited conditions.

Conventional discovery platforms use nutrient rich laboratory media when identifying drug targets and hit compounds. However, additional genes in the biosynthesis of amino acids, vitamins, and cofactors become conditionally essential to ensure survival in the host environment[1]. Namely, MAC13772, an inhibitor of BioA, is a potent compound that shows *in vivo* efficacy against priority pathogens, such as *Acinetobacter baumannii, Klebsiella pneumoniae* and *Pseudomonas aeruginosa*[66]. Performing the screen in nutrient-limited media provides the opportunity to probe such antimetabolite compounds that are otherwise inactive in rich media[49]. This dataset can also be mined to contribute insights into transcriptional responses of other groups of antibiotics in minimal media, in which the transcriptional responses may differ from nutrient rich environments. Additionally, examining the transcriptional network in a

Gram-negative bacterium informs on the effects of molecules that perturb the outer membrane. Membrane perturbants are often used in combination for their ability to potentiate the activity of other antibiotics[67]. This presents an attractive area for therapeutic opportunities; however, the exact mechanisms of action of these antibiotics, such as colistin, are still not very well understood[67]. This study provides the opportunity to gain a better understanding of the ways in which bacteria respond to membrane damage. Although transcriptomic experiments such as RNA-seq are the current gold standards for studying changes in gene expression, it is coupled with RNA turnover and typically conducted at only a few time points[23]. The screening platform presented here provides data every 5 minutes for 24 hours. A high temporal dataset allows for analysis of co-expression, which may provide a wealth of information on promoters and their associated biological processes[3]. The dataset generated from this study offers new knowledge to the field through the use of a wide range of compounds, some of which had not been included in a study previously. Ultimately, this project offers a unique opportunity to study the basic biology behind *E. coli*'s transcriptional network under chemical stress. Furthermore, these data can serve as a hypothesis generation and MOA prediction tool to help characterize unknown compounds.

Data and plots will be made available through an online searchable server, wherein scientists can search for a gene of interest, compare within and between drug classes, and investigate specific drug conditions. Researchers will be able to explore the time course data, which can inform on order of induction for genes involved in a specific pathway or genes that may be controlled by a certain regulator. Investigating genes that

are co-expressed may provide insight into co-regulation and suggest genes that may be closely related in terms of biological processes or belonging to the same stress response. This tool will contribute to existing transcriptional response knowledge and may help generate new and intriguing hypotheses.

Another interesting application of this dataset is to mine for diagnostic reporters for the detection of a specific group of chemicals. In the current drug discovery pipeline, many chemical screens will begin with an assay for growth inhibition, followed by counter screens to select for compounds of particular interest based on certain criteria. Development of selective primary screening assays that can sensitively detect compounds of interest can help avoid off-target compounds and advance through the steps of the discovery stages more rapidly[68]. In a study by Czarny et al, an autonomous luminescence gene cluster driven by the promoter of *ywaC* was used in a primary screen to detect inhibitors of cell wall biosynthesis[68]; 9 novel compounds that target cell wall biosynthesis in *Bacillus subtilis* were discovered in this screen[68]. In this study, we present *sulA* as a reporter for compounds whose primary MOA involves damaging DNA. This includes the fluoroquinolones[13] which target DNA gyrase and topoisomerase, as well as the mitomycins[69] and nitrofurans[70] which bind to DNA and form cross-links. The aminocoumarin group of antibiotics, such as novobiocin, interfere with the ATPase activity of DNA gyrase[71]. The chromosomal integrity is maintained and no double-stranded breaks occur; the SOS response is not activated in this case[71]. The *recA* reporter is complementary to the *sulA* reporter; *recA* reports on early stage SOS response while *sulA* indicates the activation of late stage SOS response[72]. Here, we observe that the cell

wall biosynthesis inhibitors ampicillin and D-cycloserine induces *recA* to a moderate extent, indicating DNA damage as a secondary effect. This result is in agreement with what is currently known about cell wall inhibitors and SOS response[73]. The *fpr* reporter is very selective for detecting superoxide producing antibiotics, including trimethoprim[46] and paraquat[74]. The results suggest that A22 may also induce oxidative stress. Lastly, the *rcsA* promoter was induced to the highest degree by the penam group of cell wall inhibitors, and less so by the cephems, as seen in the literature[10]. However, the membrane perturbants did not induce *rcsA* as expected. Indeed, we have observed, through our BARseq experiments, that many genes are more sensitive to colistin, a membrane perturbant, in minimal medium relative to nutrient rich medium (Unpublished, Singh, M). As such, we hypothesize that the stringent response may lead to changes in the membrane stress response through regulation of metabolic pathways associated with membrane assembly. Ultimately, these promoters have the potential to become selective diagnostic reporters with low background fluorescence.

As one of the main objectives of this study, we demonstrate that investigating bacterial transcriptional responses to exogenous chemicals can help determine the MOA class for an unknown compound. As seen through the linear discriminant analysis, MOA classes were separated by variations in transcriptional fingerprints and occupy unique places in the LDA 3-dimensional space. Clustering of any unknown compounds with this set of known antibiotics can help infer on the MOA class of the unknown molecule. Key promoters responsible for the distinction between classes have been identified and can be tested with unknown molecules to investigate MOA hypotheses. Although some

antibiotics may activate stress responses due to a distinct secondary effect that is not conserved within its MOA class, the differences in transcriptional signature between each class and the similarity within each class is sufficient for distinguishing between classes.

Furthermore, a neural network was built using time course fluorescence patterns from the set of known antibiotics. Deep-learning approaches have been used in MOA prediction for bioactive chemical queries, typically utilizing imaging techniques. These approaches implement high-content screening approaches to collect cell and organelle features from fluorescent stains, to train neural networks. Feature acquisition is generally limited by the fluorescent probes chosen to screen, as well as the number of response phenotypes to the chemical probes they are exposed to. This is in contrast to reporter library-based methods, which have features associated with every reporter, but also enable the collection of temporal features when acquired kinetically. The temporal dimension effectively expands the data that can be associated with a chemical structure and the deep learning model jointly identifies patterns in the data (co-expression)[32]. Where imaging methods may extract > 800 features[75] associated with a chemical structure, kinetic transcriptional reporter methods can provide even more. Namely, Zoffmann *et al*.[76] examined traditional morphological fingerprints after compound exposure in bacteria, aiming to identify MOA for compounds with unknown target. A series of 15 boronate compounds were used as test compounds and three of these compounds were found to have a high similarity score (0.66-0.72) to FabI inhibitor, triclosan[76]. However, a limitation to looking at diagnostic morphological features is that some chemicals, such as nitroxoline, will not produce a distinct morphological

fingerprint[76]. In fact, in their work, only 5 of the 15 borate analogs demonstrated a fingerprint based on the morphology defect they elicit[76]. Using transcriptional information to complement the morphology, nitroxoline resulted in a transcriptional response that is distinct from other MOA classes[76]. The authors suggest that these combined datasets can help guide hit-to-lead campaigns and can be applied to a broad range of biological problems.

In this study, the deep learning model was tested by generating new data for drugs that were either present or absent from the training set and predicting the mechanism of action using the model. While the linear discriminate analysis could report on the MOA class of unknowns based on clustering, the deep learning model may be able to identify a specific chemical in the training set that is most similar to the test compound in terms of transcriptional response. Here, we show that the model was able to accurately predict the identity of cefmetazole, polymyxin B, and cinoxacin as enoxacin, a structurally similar fluoroquinolone, to a high degree accuracy (80%-100% accuracy). Future work needs to prioritize further testing of this model for its prediction accuracy. One test method would be the "leave-one-out" approach, where each compound in the training set is removed from the group and treated as a test compound instead. This will provide a prediction output for every drug in the training set and report on the accuracy of the model with a larger sample size. An important limitation of this prediction tool is that it will not be able predict the target or MOA of a compound if the target is not already covered in the training set. In other words, if there are no known inhibitors of a target protein or the inhibitor is not included in the training set, the model will not be able to correctly identify

the target or MOA of the compound, as there is no example inhibitor for the model to learn from. Currently, the best way to combat this limitation is to continue screening compounds and cover as many protein targets as possible. Additional compounds should be screened and included in the training set; this is especially needed for MOA classes with low prediction accuracy and/or MOA classes that are underrepresented compared to other groups. As the total number of compounds in the training set increases, the layers of the neural network will be continuously adjusted and optimized to achieve a highly accurate system.

The deep learning model presented in this study was used to generate a prediction for a compound with an unknown MOA, MAC168425. The objective here was to determine whether a MOA prediction could help identify the target of an unknown compound through a series of targeted follow-up experiments. MAC168425 is an antimetabolite compound discovered by Zlitni et al. in a chemical screen looking for growth inhibition in nutrient limited medium[49]. It was observed that the activity of MAC168425 was suppressed by glycine and to a lesser extent by L-threonine[49]. The hypothesis was that the inhibitory activity of MAC168425 was mediated through the biosynthesis or utilization of glycine[49]; however, its enzymatic target remained unknown.

Here, we continued the MOA investigation on this compound; the model predicted trimethoprim for MAC168425. Overexpression of the enzymatic target of trimethoprim, FolA, does not lead to any changes in its activity, suggesting that FolA is likely not the target of MAC168425. However, it is important to note that the

overexpression of a target protein does not necessarily lead to resistance against the compound[50]. Palmer and Kishony stated that target overexpression with compounds that divert metabolic flux, such as sulfonamides and coumermycin A, will not lead to resistance to the drug[50]. Nonetheless, we reasoned that the target of MAC168425 may be either novel or is not yet covered by a compound in the training set. We proceeded with the hypothesis that if the training set does not contain an existing inhibitor of the target of MAC168425, the prediction model will compute a drug output that most closely resemble MAC168425 in terms of transcriptional response. We hypothesized that MAC168425 may still be closely related to the folate biosynthesis pathway or folate utilization.

After examining the combination effects of MAC168425 with sulfamethoxazole and trimethoprim, we reasoned that MAC168425 could be 1) targeting DHFR as is trimethoprim but using a different mechanism, 2) targeting dihydropteroate synthase (FolP), the target of sulfamethoxazole at a different site, or 3) inhibiting an earlier or later step in a common metabolic pathway. The first and second explanations seem unlikely since Zlitni et al[49] showed that the inhibitory activity of trimethoprim and sulfamethoxazole cannot be suppressed by adding glycine to the medium. In fact, Kwon et al states that in glycine-supplemented media, trimethoprim will cause thymine-less cell death[77]. These results suggest that MAC168425 may be targeting an earlier or later step in a common pathway. However, after investigating the sensitivity of folate biosynthesis deletion mutants to MAC168425, we discovered that mutants involved in the steps of the folate biosynthesis largely showed no change in sensitivity.

Another approach at discovering the enzymatic target of a molecule is to generate spontaneous resistant mutants. All three mutants contained an insertion of a mobile element, IS5, upstream of both *csrD* and *yhdH*. Through RT-qPCR and dose response assays, we determined that the IS5 upregulates *yhdH* to confer resistance against MAC168425. Although overexpression of YhdH suppresses the activity of MAC168425, it is likely not the enzyme target of MAC168425 because it is not an essential enzyme required for growth in minimal medium. As such, overexpression of YhdH may be a result of compensating for the actual inhibitory activity of MAC168425 or is co-transcribed with the target of MAC168425 by a common regulator.

Since the physiological role of YhdH in *E. coli* is poorly characterized, we studied the homolog of YhdH in the Roseobacter group of marine bacteria, named AcuI[54]. AcuI is involved the catabolism of DMSP, which is not used by *E. coli*[54]. However, it was intriguing that many *acuI*-like genes were located in close proximity with other DMSP catabolism genes such as *dmdA*[54]. Many of these *dmdA* genes were located directly downstream of *acuI* and were co-transcribed[54]; this close association prompted investigations into *E. coli* protein homologs of DmdA. To determine what the physiological role of YhdH may be in *E. coli*, we further explored *dmdA* as it is often co-regulated with *yhdH*. DmdA is a member of the aminomethyltransferase-like protein family[59]. DmdA shares structural similarity with the proteins GcvT[60] and YgfZ[61]. GcvT is a member of the glycine cleavage system responsible for the oxidation of glycine[62]. YgfZ has been implicated in the synthesis and repair of iron-sulfur clusters[63], response to oxidative stress[64], and tRNA modification[65]. Deletion mutants in *gcvT* and *ygfZ* both

showed increased sensitivity to MAC168425. Interestingly, the similarity between these proteins mostly comes from the fact they are all folate-binding or folate-dependent enzymes[61,62]. The three-dimensional structure of DmdA, GcvT, and YgfZ contains similar folate binding sites; these three enzymes all use tetrahydrofolate to accept a formaldehyde unit, yielding 5,10-methylene-THF[59,61,78]. Indeed, this interaction with glycine cleavage and folate-dependent enzymes may explain the glycine suppression phenotype we observed previously, as well as the trimethoprim prediction output from the deep learning model.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

As demonstrated in this study, *E. coli* respond to chemical stimuli in their environments. These transcriptional responses can be captured as unique and reproducible signatures and used to train deep-learning models to predict the MOA of unknown compounds. This project offers a unique opportunity to study bacterial response to chemical stress in a large-scale, kinetic, and comprehensive manner. The distinct response signatures induced by each antibiotic class can not only help researchers generate hypotheses about compounds with an unknown MOA, but also gain further insight into the off-target or secondary effects of known antibiotics. This high-throughput phenomics platform provides a means to study genome-scale transcriptional responses to a wide range of environmental stressors, which stretches beyond the limits of chemicals. Innovative methods as such are crucial in advancing the pipeline of modern drug discovery.

Future work needs to prioritize further testing of this model for its prediction accuracy. One test method would be the "leave-one-out" approach, where each compound in the training set is removed the group and treated as a test compound instead. This would provide a prediction output for every drug in the training set and report on the accuracy of the model with a larger sample size. An important limitation of this prediction tool is that it will not be able predict the target or MOA of a compound if the target is not already covered in the training set. In other words, if there are no known inhibitors of a target protein or the inhibitor is not included in the training set, the model will not be able to correctly identify the target or MOA of the compound, as there is no

example inhibitor for the model to learn from. Currently, the best way to combat this limitation will be to continue screening compounds and cover as many protein targets as possible. Additional compounds should be screened and included in the training set; this is especially needed for MOA classes with low prediction accuracy and/or MOA classes that are underrepresented compared to other groups. As the total number of compounds in the training set increases, the layers of the neural network will be continuously adjusted and optimized to achieve a highly accurate system.

Many questions still remain regarding the MOA of MAC168425. Future work is required for complete characterization. Firstly, our current hypothesis was established on the assumption that upregulation of YhdH would also lead to upregulation of DmdA, which may be GcvT and YgfZ *in E. coli*. To test this, RT-qPCR needs to be performed on the MAC168425 resistant mutants and their parent wildtype strain. Since YgfZ has been implicated as an oxidative stress response protein[64], we can further probe the interaction between YgfZ and MAC168425 through checkerboards with superoxide inducing molecules such as paraquat[45]. One confounding factor for the *ygfZ*-null strain is that previous work has reported that a *ygfZ*-null mutant grows very poorly in a nutrient limited condition[79]. Future work should prioritize generating a new *ygfZ*-null mutant with PCR confirmation. If this mutant is indeed essential in minimal media, then it could very likely be the target of MAC168425.

REFERENCES

(1) Brown, E. D.; Wright, G. D. Antibacterial Drug Discovery in the Resistance Era. *Nature*. 2016, pp 336–343.
(2) *Bacterial Stress Responses*, 2nd ed.; Storz, G., Hengge, R., Eds.; 2011.
(3) French, S.; Coutts, B. E.; Brown, E. D. Open-Source High-Throughput Phenomics of Bacterial Promoter-Reporter Strains. *Cell Syst*. **2018**, *7*, 1–8.
(4) Cashel, M.; Potrykus, K. Stringent Response. In *Brenner's Encyclopedia of Genetics: Second Edition*; 2013; pp 573–575.
(5) Cashel, M.; Kalbacher, B. The Control of Ribonucleic Acid Synthesis in Escherichia Coli. V. Characterization of a Nucleotide Associated with the Stringent Response. *J. Biol. Chem*. **1970**, *245* (9), 2309–2318.
(6) Mittenhuber, G. Comparative Genomics and Evolution of Genes Encoding Bacterial (p)PpGpp Synthetases/Hydrolases (the Rel, RelA and SpoT Proteins). *J Mol Microbiol Biotechnol* **2001**, *3* (4), 585–600.
(7) Paul, B. J.; Berkmen, M. B.; Gourse, R. L. DksA Potentiates Direct Activation of Amino Acid Promoters by PpGpp. *Proc. Natl. Acad. Sci*. **2005**, *102* (22), 7823–7828.
(8) Maxson, M. E.; Darwin, A. J. Identification of Inducers of the Yersinia Enterocolitica Phage Shock Protein System and Comparison to the Regulation of the RpoE and Cpx Extracytoplasmic Stress Responses. *J. Bacteriol*. **2004**, *186* (13), 4199–4208.
(9) Darwin, A. J. The Phage-Shock-Protein Response. *Molecular Microbiology*. 2005, pp 621–628.
(10) Laubacher, M. E.; Ades, S. E. The Rcs Phosphorelay Is a Cell Envelope Stress Response Activated by Peptidoglycan Stress and Contributes to Intrinsic Antibiotic Resistance. *J. Bacteriol*. **2008**, *190* (6), 2065–2074.
(11) Klein, G.; Lindner, B.; Brabetz, W.; Brade, H.; Raina, S. Escherichia Coli K-12 Suppressor-Free Mutants Lacking Early Glycosyltransferases and Late Acyltransferases. Minimal Lipopolysaccharide Structure and Induction of Envelope Stress Response. *J. Biol. Chem*. **2009**, *284* (23), 15369–15389.
(12) De Las Penas, A.; Connolly, L.; Gross, C. A. σ(E) Is an Essential Sigma Factor in Escherichia Coli. *J. Bacteriol*. **1997**, *179* (21), 6862–6864.
(13) Qin, T.-T.; Kang, H.-Q.; Ma, P.; Li, P.-P.; Huang, L.-Y.; Gu, B. SOS Response and Its Regulation on the Fluoroquinolone Resistance. *Ann Transl Med* **2015**, *3* (22), 1–17.
(14) Aldred, K. J.; Kerns, R. J.; Osheroff, N. Mechanism of Quinolone Action and Resistance. *Biochemistry*. 2014, pp 1565–1574.
(15) Phillips, I.; Culebras, E.; Moreno, F.; Baquero, F. Induction of the SOS Response by New 4-Quinolones. *J. Antimicrob. Chemother*. **1987**, *20* (5), 631–638.
(16) Keller, K. L.; Overbeck-Carrick, T. L.; Beck, D. J. Survival and Induction of SOS in Escherichia Coli Treated with Cisplatin, UV-Irradiation, or Mitomycin C Are Dependent on the Function of the RecBC and RecFOR Pathways of Homologous Recombination. *Mutat. Res. - DNA Repair* **2001**, *486* (1), 21–29.

(17) Iwasaki, H.; Nakata, A.; Walker, G. C.; Shinagawa, H. The Escherichia Coli PolB Gene, Which Encodes DNA Polymerase II, Is Regulated by the SOS System. *J. Bacteriol.* **1990**, *172* (11), 6268–6273.
(18) Cordell, S. C.; Robinson, E. J. H.; Lowe, J. Crystal Structure of the SOS Cell Division Inhibitor SulA and in Complex with FtsZ. *Proc. Natl. Acad. Sci.* **2003**, *100* (13), 7889–7894.
(19) Power, E. G. M.; Phillips, I. Induction of the SOS Gene (UmuC) by 4-Quinolone Antibacterial Drugs. *J. Med. Microbiol.* **1992**, *36* (2), 78–82.
(20) Cohen, S. E.; Foti, J. J.; Simmons, L. A.; Walker, G. C. The SOS Regulatory Network. *EcoSal Plus* **2014**, *3* (1).
(21) Heath, R. J.; Yu, Y. T.; Shapiro, M. A.; Olson, E.; Rock, C. O. Broad Spectrum Antimicrobial Biocides Target the FabI Component of Fatty Acid Synthesis. *J. Biol. Chem.* **1998**.
(22) Price, A. C.; Choi, K. H.; Heath, R. J.; Li, Z.; White, S. W.; Rock, C. O. Inhibition of β-Ketoacyl-Acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin: Structure and Mechanism. *J. Biol. Chem.* **2001**.
(23) Hutter, B.; Schaab, C.; Albrecht, S.; Borgmann, M.; Brunner, N. A.; Freiberg, C.; Ziegelbauer, K.; Rock, C. O.; Ivanov, I.; Loferer, H. Prediction of Mechanisms of Action of Antibacterial Compounds by Gene Expression Profiling. *Antimicrob. Agents Chemother.* **2004**, *48* (8), 2838–2844.
(24) Mitosch, K.; Rieckh, G.; Bollenbach, T. Noisy Response to Antibiotic Stress Predicts Subsequent Single-Cell Survival in an Acidic Environment. *Cell Syst.* **2017**, *4* (4), 393-403.e5.
(25) Hughes, T. R.; Jones, A. R.; Bennett, H. A.; Marton, M. J.; King, A. M.; Roberts, C. J.; Stoughton, R.; Armour, C. D.; Coffey, E.; Dai, H.; et al. Functional Discovery via a Compendium of Expression Profiles. *Cell* **2000**, *102* (1), 109–126.
(26) Bandow, J. E.; Brötz, H.; Leichert, L. I. O.; Labischinski, H.; Hecker, M. Proteomic Approach to Understanding Antibiotic Action. *Antimicrob. Agents Chemother.* **2003**, *47* (3), 948–955.
(27) Goh, E.-B.; Yim, G.; Tsui, W.; McClure, J.; Surette, M. G.; Davies, J. Transcriptional Modulation of Bacterial Gene Expression by Subinhibitory Concentrations of Antibiotics. *Proc. Natl. Acad. Sci.* **2002**, *99* (26), 17025–17030.
(28) Utaida, S.; Dunman, P. M.; Macapagal, D.; Murphy, E.; Projan, S. J.; Singh, V. K.; Jayaswal, R. K.; Wilkinson, B. J. Genome-Wide Transcriptional Profiling of the Response of Staphylococcus Aureus to Cell-Wall-Active Antibiotics Reveals a Cell-Wall-Stress Stimulon. *Microbiology* **2003**.
(29) Hutter, B.; Fischer, C.; Jacobi, A.; Schaab, C.; Loferer, H. Panel of Bacillus Subtilis Reporter Strains Indicative of Various Modes of Action. *Antimicrob. Agents Chemother.* **2004**.
(30) O'Rourke, A.; Beyhan, S.; Choi, Y.; Morales, P.; Chan, A. P.; Espinoza, J. L.; Dupont, C. L.; Meyer, K. J.; Spoering, A.; Lewis, K.; et al. Mechanism-Of-Action Classification of Antibiotics by Global Transcriptome Profiling. *Antimicrob. Agents Chemother.* **2020**.
(31) Zaslaver, A.; Bren, A.; Ronen, M.; Itzkovitz, S.; Kikoin, I.; Shavit, S.;

Liebermeister, W.; Surette, M. G.; Alon, U. A Comprehensive Library of Fluorescent Transcriptional Reporters for Escherichia Coli. *Nat. Methods* **2006**, *3* (8), 623–628.

(32) French, S.; Coutts, B. E.; Brown, E. D. Open-Source High-Throughput Phenomics of Bacterial Promoter-Reporter Strains. *Cell Syst.* **2018**, *7* (3), 339-346.e3.

(33) French, S.; Guo, A. B. Y.; Brown, E. D. A Comprehensive Guide to Dynamic Analysis of Microbial Gene Expression Using the 3D-Printed PFIbox and a Fluorescent Reporter Library. *Nat. Protoc.* **2020**, *15* (575–603).

(34) French, S.; Mangat, C.; Bharat, A.; Côté, J.-P.; Mori, H.; Brown, E. D.; Bharat, A.; Côté, J.-P.; Mori, H.; Mangat, C.; et al. A Robust Platform for Chemical Genomics in Bacterial Systems. *Mol. Biol. Cell* **2016**, *27* (6), 1015–1025.

(35) Hartig, S. M. Basic Image Analysis and Manipulation in ImageJ. *Curr. Protoc. Mol. Biol.* **2013**, No. SUPPL.102.

(36) Karp, P. D. The EcoCyc Database. *Nucleic Acids Res.* **2002**, *30* (1), 56–58.

(37) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection. *Mol. Syst. Biol.* **2006**.

(38) Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H. Complete Set of ORF Clones of Escherichia Coli ASKA Library (A Complete Set of E. Coli K-12 ORF Archive): Unique Resources for Biological Research. *DNA Res.* **2005**.

(39) Datta, S.; Costantino, N.; Court, D. L. A Set of Recombineering Plasmids for Gram-Negative Bacteria. *Gene* **2006**.

(40) Pfaffl, M. W. A New Mathematical Model for Relative Quantification in Real-Time RT-PCR. *Nucleic Acids Res.* **2001**.

(41) French, S.; Mangat, C.; Bharat, A.; Côté, J.-P.; Mori, H.; Brown, E. D. A Robust Platform for Chemical Genomics in Bacterial Systems. *Mol. Biol. Cell* **2016**, *27* (6), 1015–1025.

(42) Battesti, A.; Majdalani, N.; Gottesman, S. The RpoS-Mediated General Stress Response in Escherichia Coli . *Annu. Rev. Microbiol.* **2010**, *65* (1), 189–213.

(43) Layton, J. C.; Foster, P. L. Error-Prone DNA Polymerase IV Is Controlled by the Stress-Response Sigma Factor, RpoS, in Escherichia Coli. *Mol. Microbiol.* **2003**, *50* (2), 549–561.

(44) Jair, K. W.; Fawcett, W. P.; Fujita, N.; Ishihama, A.; Wolf, R. E. Ambidextrous Transcriptional Activation by SoxS: Requirement for the C-Terminal Domain of the RNA Polymerase Alpha Subunit in a Subset of Escherichia Coli Superoxide-Inducible Genes. *Mol. Microbiol.* **1996**.

(45) Liochev, S. I.; Hausladen, A.; Beyer, W. F.; Fridovich, I. NADPH:Ferredoxin Oxidoreductase Acts as a Paraquat Diaphorase and Is a Member of the SoxRS Regulon. *Proc. Natl. Acad. Sci. U. S. A.* **1994**.

(46) Giroux, X.; Su, W. L.; Bredeche, M. F.; Matic, I. Maladaptive DNA Repair Is the Ultimate Contributor to the Death of Trimethoprim-Treated Cells under Aerobic and Anaerobic Conditions. *Proc. Natl. Acad. Sci. U. S. A.* **2017**.

(47) Stout, V.; Torres-Cabassa, A.; Maurizi, M. R.; Gutnick, D.; Gottesman, S. RcsA, an Unstable Positive Regulator of Capsular Polysaccharide Synthesis. *J. Bacteriol.* **1991**.
(48) Farris, C.; Sanowar, S.; Bader, M. W.; Pfuetzner, R.; Miller, S. I. Antimicrobial Peptides Activate the Rcs Regulon through the Outer Membrane Lipoprotein RcsF. *J. Bacteriol.* **2010**.
(49) Zlitni, S.; Ferruccio, L. F.; Brown, E. D. Metabolic Suppression Identifies New Antibacterial Inhibitors under Nutrient Limitation. *Nat. Chem. Biol.* **2013**, *9* (12), 796–804.
(50) Palmer, A. C.; Kishony, R. Opposing Effects of Target Overexpression Reveal Drug Mechanisms. *Nat. Commun.* **2014**.
(51) Minato, Y.; Dawadi, S.; Kordus, S. L.; Sivanandam, A.; Aldrich, C. C.; Baughn, A. D. Mutual Potentiation Drives Synergy between Trimethoprim and Sulfamethoxazole. *Nat. Commun.* **2018**, *9* (1).
(52) Deatherage, D. E.; Barrick, J. E. Identification of Mutations in Laboratory-Evolved Microbes from next-Generation Sequencing Data Using Breseq. *Methods Mol. Biol.* **2014**.
(53) Jonas, K.; Tomenius, H.; Römling, U.; Georgellis, D.; Melefors, Ö. Identification of YhdA as a Regulator of the Escherichia Coli Carbon Storage Regulation System. *FEMS Microbiol. Lett.* **2006**.
(54) Todd, J. D.; Curson, A. R. J.; Sullivan, M. J.; Kirkwood, M.; Johnston, A. W. B. The Ruegeria Pomeroyi Acui Gene Has a Role in DMSP Catabolism and Resembles YhdH of E. Coli and Other Bacteria in Conferring Resistance to Acrylate. *PLoS One* **2012**.
(55) Engler, J. A.; van Bree, M. P. The Nucleotide Sequence and Protein-Coding Capability of the Transposable Element IS5. *Gene* **1981**.
(56) Sawers, R. G. Expression of Fnr Is Constrained by an Upstream IS5 Insertion in Certain Escherichia Coli K-12 Strains. *J. Bacteriol.* **2005**.
(57) Zhang, Z.; Yen, M. R.; Saier, M. H. Precise Excision of IS5 from the Intergenic Region between the FucPIK and the FucAO Operons and Mutational Control of FucPIK Operon Expression in Escherichia Coli. *J. Bacteriol.* **2010**.
(58) Schnetz, K.; Rak, B. IS5: A Mobile Enhancer of Transcription in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **1992**.
(59) Schuller, D. J.; Reisch, C. R.; Moran, M. A.; Whitman, W. B.; Lanzilotta, W. N. Structures of Dimethylsulfoniopropionate-Dependent Demethylase from the Marine Organism Pelagabacter Ubique. *Protein Sci.* **2012**.
(60) Okamura-Ikeda, K.; Hosaka, H.; Maita, N.; Fujiwara, K.; Yoshizawa, A. C.; Nakagawa, A.; Taniguchi, H. Crystal Structure of Aminomethyltransferase in Complex with Dihydrolipoyl-H-Protein of the Glycine Cleavage System: Implications for Recognition of Lipoyl Protein Substrate, Disease-Related Mutations, and Reaction Mechanism. *J. Biol. Chem.* **2010**.
(61) Teplyakov, A.; Obmolova, G.; Sarikaya, E.; Pullalarevu, S.; Krajewski, W.; Galkin, A.; Howard, A. J.; Herzberg, O.; Gilliland, G. L. Crystal Structure of the YgfZ Protein from Escherichia Coli Suggests a Folate-Dependent Regulatory Role

in One-Carbon Metabolism. *J. Bacteriol.* **2004**.

(62) OKAMURA-IKEDA, K.; OHMURA, Y.; FUJIWARA, K.; MOTOKAWA, Y. Cloning and Nucleotide Sequence of the Gcv Operon Encoding the Escherichia Coli Glycine-cleavage System. *Eur. J. Biochem.* **1993**.

(63) Waller, J. C.; Ellens, K. W.; Hasnain, G.; Alvarez, S.; Rocca, J. R.; Hanson, A. D. Evidence That the Folate-Dependent Proteins YgfZ and MnmEG Have Opposing Effects on Growth and on Activity of the Iron-Sulfur Enzyme MiaB. *J. Bacteriol.* **2012**.

(64) Lin, C. N.; Syu, W. J.; Sun, W. S. W.; Chen, J. W.; Chen, T. H.; Don, M. J.; Wang, S. H. A Role of YgfZ in the Escherichia Coli Response to Plumbagin Challenge. *J. Biomed. Sci.* **2010**.

(65) Ote, T.; Hashimoto, M.; Ikeuchi, Y.; Su'etsugu, M.; Suzuki, T.; Katayama, T.; Kato, J. I. Involvement of the Escherichia Coli Folate-Binding Protein YgfZ in RNA Modification and Regulation of Chromosomal Replication Initiation. *Mol. Microbiol.* **2006**.

(66) Carfrae, L. A.; MacNair, C. R.; Brown, C. M.; N, T. C.; Weber, B. S.; Zlitni, S.; Rao, V. N.; Chun, J.; Junop, M. S.; Coombes, B. K.; et al. Biotin Biosynthesis Is an Overlooked Antibiotic Target. *Nat. Microbiol.* **2020**, *5* (93–101).

(67) MacNair, C. R.; Stokes, J. M.; Carfrae, L. A.; Fiebig-Comyn, A. A.; Coombes, B. K.; Mulvey, M. R.; Brown, E. D. Overcoming Mcr-1 Mediated Colistin Resistance with Colistin in Combination with Other Antibiotics. *Nat. Commun.* **2018**.

(68) Czarny, T. L.; Perri, A. L.; French, S.; Brown, E. D. Discovery of Novel Cell Wall-Active Compounds Using Pywac, a Sensitive Reporter of Cell Wall Stress, in the Model Gram-Positive Bacterium Bacillus Subtilis. *Antimicrob. Agents Chemother.* **2014**.

(69) Sekiguchi, M.; Takagi, Y. Effect of Mitomycin C on the Synthesis of Bacterial and Viral Deoxyribonucleic Acid. *BBA - Biochim. Biophys. Acta* **1960**.

(70) Mc Osker, C. C.; Fitzpatrick, P. M. Nitrofurantoin: Mechanism of Action and Implications for Resistance Development in Common Uropathogens. *J. Antimicrob. Chemother.* **1994**.

(71) Jara, L. M.; Pérez-Varela, M.; Corral, J.; Arch, M.; Cortés, P.; Bou, G.; Aranda, J.; Barbé, J. Novobiocin Inhibits the Antimicrobial Resistance Acquired through DNA Damage-Induced Mutagenesis in Acinetobacter Baumannii. *Antimicrob. Agents Chemother.* **2016**.

(72) Maslowska, K. H.; Makiela-Dzbenska, K.; Fijalkowska, I. J. The SOS System: A Complex and Tightly Regulated Response to DNA Damage. *Environmental and Molecular Mutagenesis*. 2019.

(73) Maiques, E.; Úbeda, C.; Campoy, S.; Salvador, N.; Lasa, Í.; Novick, R. P.; Barbé, J.; Penadés, J. R. β-Lactam Antibiotics Induce the SOS Response and Horizontal Transfer of Virulence Factors in Staphylococcus Aureus. *J. Bacteriol.* **2006**.

(74) Hassan, H. M.; Fridovich, I. Superoxide Radical and the Oxygen Enhancement of the Toxicity of Paraquat in Escherichia Coli. *J. Biol. Chem.* **1978**.

(75) Simm, J.; Klambauer, G.; Arany, A.; Steijaert, M.; Wegner, J. K.; Gustin, E.; Chupakhin, V.; Chong, Y. T.; Vialard, J.; Buijnsters, P.; et al. Repurposing High-

Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. *Cell Chem. Biol.* **2018**.

(76) Zoffmann, S.; Vercruysse, M.; Benmansour, F.; Maunz, A.; Wolf, L.; Blum Marti, R.; Heckel, T.; Ding, H.; Truong, H. H.; Prummer, M.; et al. Machine Learning-Powered Antibiotics Phenotypic Drug Discovery. *Sci. Rep.* **2019**, *9* (1), 1–14.

(77) Kwon, Y. K.; Higgins, M. B.; Rabinowitz, J. D. Antifolate-Induced Depletion of Intracellular Glycine and Purines Inhibits Thymineless Death in E. Coli. *ACS Chem. Biol.* **2010**, *5* (8), 787–795.

(78) Okamura-Ikeda, K.; Fujiwara, K.; Motokawa, Y. Identification of the Folate Binding Sites on the Escherichia Coli T- Protein of the Glycine Cleavage System. *J. Biol. Chem.* **1999**.

(79) Waller, J. C.; Alvarez, S.; Naponelli, V.; Lara-Nuñez, A.; Blaby, I. K.; Da Silva, V.; Ziemak, M. J.; Vickers, T. J.; Beverley, S. M.; Edison, A. S.; et al. A Role for Tetrahydrofolates in the Metabolism of Iron-Sulfur Clusters in All Domains of Life. *Proc. Natl. Acad. Sci. U. S. A.* **2010**.