Engineering Bacterial Gene Expression:
Applications towards Biofuels and Antibiotic Resistance

By:

Peter Britton Otoupal

Bachelor of Science, Chem. Eng. Honors, the University of Texas at Austin, 2013

Master of Science, Chem. and Bio. Eng., the University of Colorado at Boulder, 2015

A thesis submitted to the

Faculty of the Graduate School of the

The University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Chemical and Biological Engineering

Advisor: Dr. Anushree Chatterjee

2018

This thesis entitled:

Engineering Bacterial Gene Expression:

Applications towards Biofuels and Antibiotic Resistance

written by Peter Britton Otoupal

has been approved for the Department of Chemical and Biological Engineering by:

_____ _____

Dr. Anushree Chatterjee, Committee Chair Date

_____ _____

Dr. _____, Committee Member Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

ABSTRACT

Otoupal, Peter Britton (Ph.D., Chemical and Biological Engineering)

Engineering bacterial gene expression: applications towards biofuels and antibiotic resistance

Thesis directed by Prof. Anushree Chatterjee

Nature is rarely static. Organisms live in diverse and stressful environments that necessitate rapid response strategies for survival. Microorganisms have responded to this by evolving bet-hedging, wherein they exhibit constitutively heterogeneous gene expression to maximize fitness across numerous background. The goal of this thesis is to “hijack” this phenomenon using novel gene expression engineering techniques to alter how bacteria respond to their environments, in order to address pressing societal concerns.

This begins with a systematic exploration of how bacterial gene expression naturally responds to antibiotics and biofuels. This reveals promising gene candidates for targeted manipulation, for which a library of CRISPR gene expression perturbation devices is constructed. This library is applied to *Escherichia coli* during exposure to antibiotics and biofuels, and the impact of CRISPR perturbation on growth and fitness is quantified. Many perturbations show significant non-heritable improvements or detriments on growth, indicating the potential of this approach for biofuel and antibiotic applications. To improve the desired bacterial response, individual perturbations are combined in a multiplexed fashion. A significant trend towards lower fitness as more perturbations are combined emerges, which is supported by a systematic exploration of

combinatorial perturbation libraries. This trend is correlated to a diminished adaptive potential, suggesting the applicability of multiplexed perturbations for restricting bacterial evolution. In a similar vein, the ability of gene expression perturbations to synergize with antibiotics is explored to identify novel potentiating therapies. Significant gene-drug synergies are characterized and used to potentiate antibiotic treatment in an infection model. Therapeutic peptide nucleic acid molecules are subsequently designed to re-sensitize clinically isolated multidrug-resistant bacteria to treatment. Finally, this thesis expands our available synthetic biology toolkit for manipulating gene expression by outlining novel CRISPR engineering strategies. Deactivated CRISPR proteins are fused with bacterial initiation factor one, and the potential for these constructs to increase translation rates by promoting 30S subunit binding to the ribosome is explored. The design of a smart antibiotic utilizing a CRISPR-holin RNA-based kill switch is also presented. Collectively, this thesis demonstrates the power that manipulating gene expression has in affecting desired phenotypes in bacteria.

ACKNOWLEDGMENTS

This work presented here would be impossible without the support of my advisor, Dr. Anushree Chatterjee. I cannot possibly convey how influential she has been throughout my graduate studies, and indeed the trajectory of my academic career. From the beginning, she has been there to motivate, encourage, and excite me in the pursuit of science. The entire time, Anushree has gone above and beyond the requirements of an advisor, and exuded an infectious passion for research that inspired me to accomplish things I never thought would be possible. No one has helped build my confidence as a scientist as much as she has. I will forever treasure the hours of mentorship she provided and hope our friendship lasts for years to come.

I extend a sincere thanks to my committee members: Dr. Ryan Gill, Dr. Michael Shirts, Dr. Amy Palmer, and Dr. Robert Batey. Each of these professors leads highly successful labs that demand immense amounts of time, yet none hesitated to join my committee. They have provided deeply insightful feedback that has honed my research into the best form it could be. Thank you Ryan for watching over me since my preliminary examination. The fist-bump he gave me after I passed my comprehensive exam is one of my proudest memories. Thank you Amy for taking the time to review my epistasis manuscript and pointing out flaws that greatly improved it. Thank you Rob for allowing me to perform SHAPE in his lab, and for our personal meeting where you provided invaluable advice on RNA.

Thank you also to Dr. Janet deGrazia and Dr. Andrew Goodwin, both of whom I was a teaching assistant for. I learned a great deal from each of them on effective

teaching styles, and became a better mentor because of your guidance. Most importantly, they've each taught me a great deal on how to successfully manage difficult situations, and get the best out of a team.

My passion for research began long before engaging in the work presented in this Thesis, and I must acknowledge my past mentors who cultivated it so strongly. Foremost is Dr. Hal Alper from the University of Texas, who was an incredible professor to perform undergraduate research for. He took a genuine interest in my success and set me on the track I am on today. Thanks to Dr. Jeffrey Barrick, also from the University of Texas, who taught me most of the synthetic biology I know today. His revival of the Austin iGEM team was crucial to the development of my cloning skills which I still use to this day. Thank you to Dr. Lieve Laurens of the National Renewable Energy Lab, who showed me how research can lead to a career in exciting places. And thank you to my very first PI, Dr. David Johnson of M.D. Anderson Science Park, who started me on this crazy ride when he let me intern in his lab during high school.

The Chatterjee group consists of many amazing individuals who I owe immeasurable thanks to. First and foremost is Dr. Keesha Erickson, who mentored me during my first year in graduate school and was instrumental in the development of this thesis. Her work was the impetus for virtually everything presented here, and the guidance and friendship she provided over the years was invaluable. The majority of data presented in Chapter 3 was collected and analyzed by Keesha, who deserves great credit for the chapter's existence. She was also of significant help in Chapters 5 and 7, and was the progenitor of the latter's work. Dr. Colleen Courtney was also an outstanding source of inspiration, and perhaps the best role model a graduate student could ask for.

Furthermore, her work laid the foundation for all PNA work presented in this study. Antoni Escalas-Bordoy joined the same year as me, and was a close friend who shared so many twists and turns with me on our journey. Toni was also of significant help collecting FACS data in Chapter 5. I thank the newer students in our lab, Kristen Eller, Thomas Aunins, and Jocelyn Campos for their friendship at the end of this journey, and for continuing the work after I leave. They each helped in their own capacity in the construction of Chapter 7.

Throughout the pursuit of this thesis, I have collaborated closely with Dr. Prashant Nagpal and many members of his lab. I thank him for teaching me so much of a complicated physics topic, and for including me in his exciting pursuit of new sequencing technologies. Thank you specifically to Lee Korshoj and Dr. Gary Abel for their work in our quest to develop a new platform for RNA sequencing.

This work would also have been impossible without the many undergraduate students who worked for me over the years. Madeline Sitton has worked with me as an undergraduate researcher for virtually my entire Ph.D. since the spring of 2015. By the end of her freshman year, Maddie was already a highly beneficial asset to my research. Maddie has demonstrated perhaps the most determined will and motivation that I have ever seen from an undergraduate. Despite substantial setbacks, Maddie has never once wavered in her pursuit of the science. I have never witnessed someone with such dedication and passion, and I could not be more proud of her. The initiation factor work presented in Chapter 9 was adapted from her Senior Thesis, for which she performed outstandingly during many late Saturdays in the lab. She was also of significant help in

all the cloning performed in Chapters 5-8. I know Maddie will be supremely successful at Duke University pursuing her own Ph.D.

Another undergraduate student who was invaluable to my work is William Cordell. Will has demonstrated a remarkable aptitude for research and has been an invaluable asset in my research. He is a supremely dependable undergraduate who has clearly and full-heartedly pursued a career in research. Will has especially gone above and beyond in pursuit of his Senior Thesis, which he developed largely independently. Through many long nights and weekends, Will has accomplished graduate-level research that laid the foundation for Holin work presented in Chapter 8. In tandem with Maddie, Will was of significant help in all the cloning performed in chapters 6, 7, and 9. Will has an incredibly bright future pursuing his own Ph.D. at Wisconsin, and I wish him the best.

I also thank two other undergraduates, Vismaya Bachu and Logan Collins. Vismaya was instrumental in many cloning experiments, and an incredibly bright student to mentor. Her assistance was a great boon, and I know she will go on to do great things. Logan was one of the first undergraduates I ever had the pleasure of mentoring and is one of the most intelligent people I have ever met. Despite many setbacks, he has pursued his own work doggedly and has incredible potential as a scientist.

I give thanks to my family for making me the person I am today. My parents, Mark and Nan Otoupal, overcame many significant hardships to provide the best childhood a kid could ask for. Their constant praise and encouragement as I grew up convinced me to take the leap and pursue engineering, a field that I knew absolutely no one in. They have supported me immensely throughout my Ph.D., and I could not feel their love and pride more. I thank my sister, Jessica, who was always there to brighten my day with her

bountiful cheer and love for life. My grandmother, Barbara Harris, was one of the first people to instill a love of science in me by sharing exciting scientific stories in my childhood. And thanks to my grandfather, Stewart Otoupal, for raising such an incredible family.

Finally, this work would not have been possible without our sources of funding. Thank you to the National Science Foundation for supporting me through the Graduate Research Fellowship Program. Thank you for grants from the National Science Foundation, the Defense Advanced Research Projects Agency, the William M. Keck Foundation, and the University of Colorado startup funds to Anushree.

Table of Contents

List of Tables

Main Tables

Supplementary Tables

List of Figures

Main Figures

Supplementary Figures

Chapter 1: Introduction

1.1 Gene expression: an underappreciated biological lever

Roughly twenty years ago, the world has entered into a new age: the synthetic biology era. Synthetic biology exists at the intersection of engineering and biology, with the goal of rationally programming biological systems for desired applications. The modification of gene expression has been a part of the synthetic biology toolkit since the beginning. Arguably the first demonstration of a synthetic biology application was the development of a synthetic oscillatory network based on transcriptional regulators, in which gene expression was engineered to fluctuate and periodically produce green fluorescent protein[1–3]. Despite this, synthetic biologists rarely discuss their work in terms of modulating how genes are expressed. Instead, their focus has been primarily on DNA and how it can be manipulated to engineer desired phenotypes.

This view of synthetic biology is flawed. DNA is only one step in the process of life, and expanding our scope beyond this will be essential if we are to achieve all that the field has to offer. Naturally-existing multistep pathways often rely on transcription factors and post-transcriptional processing for controlling gene function[4]. The integration of heterologous pathways into an organism often entails optimization of gene expression to ensure proper behavior of the engineered system[5]. And until the advent of Clustered Regularly Interspace Short Palindromic Repeat (CRISPR) technology, there were very few tools for directly affecting gene expression without modifying the DNA level and even fewer examples of their application[6].

Turning the focus in synthetic biology to the engineering of gene expression would have powerful implications for our ability to address a number of prominent societal concerns. For instance, this could afford better ways of engineering microbes for improved tolerance towards toxic biofuels[7,8] and other bioproducts[9–12], which would provide significant economic benefit and reduce our dependence on fossil fuels. Additionally, as the threat of rampant antibiotic resistance continues to loom menacingly[13–15], the selective engineering of gene expression to reduce bacterial fitness could couple powerfully with current antimicrobial strategies.

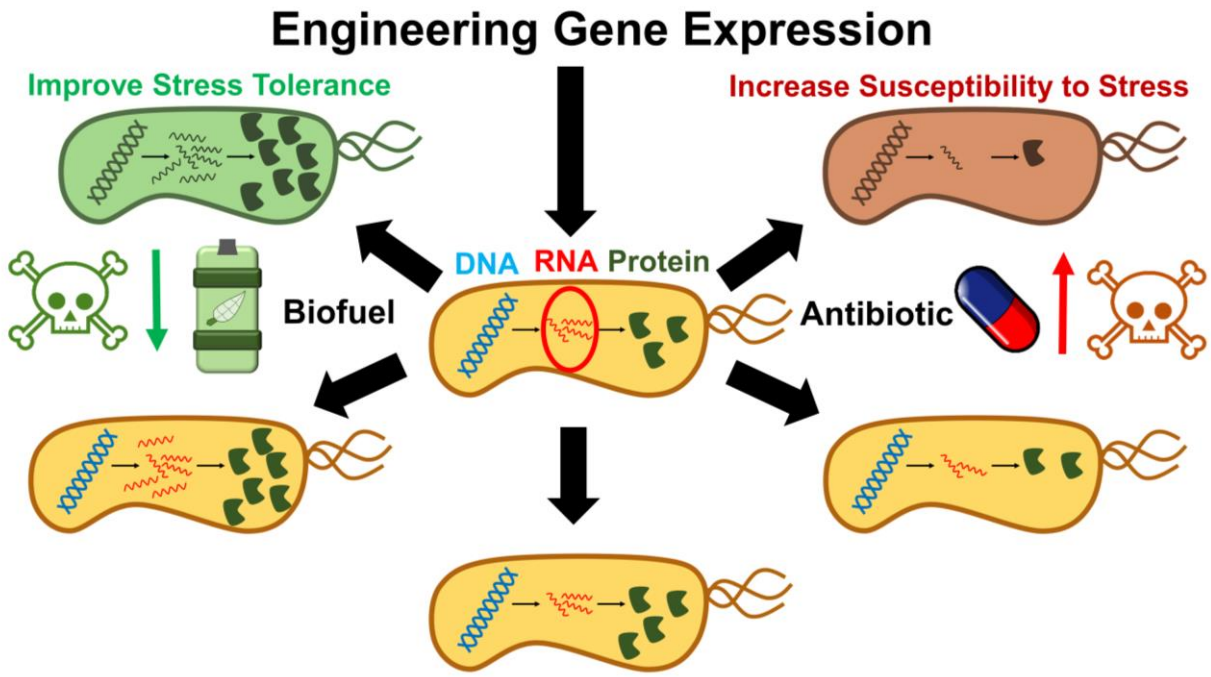


Figure 1.1 Scope of Thesis

In this thesis, I describe my attempts to take advantage of the stochastic processes that lead to gene expression heterogeneity in order to engineer bacterial phenotypes. I explain how manipulating transcription and translation enables control over the bacterial response to stress. For instance, transcription of a particular gene providing antibiotic tolerance can be knocked down to increase the bacteria's susceptibility to treatment (top right). Conversely, transcription of a different gene related to biofuel tolerance can be increased to improve bacterial growth during biofuel production (top left).

This thesis is an attempt to raise awareness for the vast untapped potential in applying synthetic biology tools to the engineering of specific gene expression patterns for desired purposes. In the following chapters, I highlight the untapped potential of this biological lever by demonstrating its applications for significantly improving bacterial biofuel tolerance, and reducing bacterial antibiotic resistance. Finally, I outline novel synthetic biology approaches I have developed for modulating gene expression, showing still has room to grow through the incorporation of new tools into the synthetic biology toolkit.

1.2 Thesis scope and organization

In this thesis, I investigate the use of novel synthetic biology technologies to rationally engineer gene expression levels in various bacterial organisms and characterize their impact on bacterial phenotypes in antimicrobial and biofuel contexts. I utilize two technologies, one based on deactivated CRISPR-associated enzymes 9 and 13 (dCas9 and dCas13), and the other based on sequence-specific RNA sequestering peptide nucleic acids (PNAs).

I begin in Chapter 2 by describing with an overview of how genes are naturally expressed in bacteria, particularly focusing on expression in *E. coli*. I then explain how CRISPR technologies came to be, and how they have specifically been applied for the manipulation of gene expression in microorganisms. I specifically focus on how dCas9, dCas9-ω, and dCas13 are used to affect gene expression. I conclude this chapter by outlining an alternative approach to modulating gene expression using PNAs.

I next outline the first works undertaken during my Ph.D. in collaboration with Dr. Keesha Erickson (Chapter 3, Erickson et. al., 2015[16] and 2017[17]), in which we systematically explored how bacteria gene expression naturally responds to exposure to a variety of antibiotic contexts. We experimentally determined *Escherichia coli* gene expression signatures that exhibited significant differences during stress exposure, revealing promising genes to manipulate for imparting desired phenotypes.

I employ the knowledge generated in these works in Chapters 4 and 5, in which I explore the use of CRISPR dCas9 and dCas9-ω gene expression perturbation technology to improve bacteria fitness during exposure to toxic concentrations of biofuels n-butanol and n-hexane (Chapter 4, Otoupal et. al., *in review*[18]), and to reduce bacteria fitness during exposure to five common antimicrobial treatments (Chapter 5, Otoupal et. al. 2016[19]).

During my investigation of CRISPR perturbations during antimicrobial exposure, I observed an intriguing phenomenon. I noticed that as multiple CRISPR perturbations were combined together in one cell, the cell's observed fitness was lower than expected, based on how the individual perturbations impacted cell growth. This led me to more carefully explore the impacts of multiplexing perturbations of separate genes on bacterial fitness. In Chapter 6 (Otoupal et. al., *in review*[20]), I describe the results of a study in which I constructed two combinatorial CRISPR libraries to systematically perturb expression one, two, three or four genes at a time. These libraries reveal a striking trend of negative epistasis (i.e. diminished fitness) as we combined perturbations; the more perturbations are introduced into a cell, the worse the bacteria grows.

I take the lessons learned from this study in Chapter 7 (Otoupal et. al., in preparation[21]), in which I attempt to develop a more robust antibiotic treatment strategy based on multiplexed gene expression perturbations. This begins with a meticulous screening of 30 gene knockouts for their ability to potentiate the activity of nine independent antibiotic treatments. Significant gene-drug synergies in which antibiotic activity was improved by gene knockouts were translated into gene knockdowns using CRISPR gene expression interference, which I demonstrate to largely replicate their knockouts' phenotypes. I apply these CRISPR interference constructs to an infection model context, demonstrating their potential for amplifying antibiotic activity in intracellular models. Finally, I design PNAs to apply gene knockdown of the most promising gene-drug synergies to five clinically isolated, multidrug-resistant bacteria. I show that these PNAs resensitize the bacteria to antibiotic treatment they were once resistant to, outlining a new paradigm for antibiotics based on gene expression perturbations.

The techniques I utilize up to this point in this thesis are largely based on dCas9 and PNAs. However, each of these techniques has their own unique drawbacks, and new tools to manipulate gene expression would expand our ability to engineer it for desired purposes. I explore the development of such tools in Chapters 8 and 9, and to an extent at the end of Chapter 4 in which I develop a novel hyper-mutator system for rapidly mutating CRISPR guide plasmids for applications in directed evolution. In Chapter 8 (Otoupal et. al., *in preparation*[22]), I reveal another novel CRISPR fusion protein based on Cas13 (the non-deactivated version) that functions as a "Smart" antibiotic. Cas13 fused to the protein degradation tag is designed to constitutively degrade mRNA coding for Holin, a toxic cell-wall degrading protein. In Chapter 9 (Otoupal et. al., *in preparation*[23]), I

outline my work in applying CRISPR dCas9 and dCas13 to control the rate of translation in *E. coli*, attempting to both decrease and increase the rate of its occurrence. I demonstrate promising results of novel CRISPR fusion proteins with the native *E. coli* gene *infA* encoding initiation factor 1 (IF1) for these purposes, and postulate further work that can be undertaken to improve this system. Collectively, the inventions outlined in these chapters demonstrate the potential for new CRISPR systems to control gene expression.

The sum of the work presented in this thesis provides evidence for great, untapped potential in applying gene expression engineering to modify bacteria for desired purposes. This is primarily explored through the lenses of applications in novel antimicrobial strategies and in improving bacterial growth in biofuels. In Chapter 10, I summarize the key findings and conclusions of this thesis and future directions for this line of research.

Chapter 2: Background

2.1 An overview of bacterial gene expression

To understand how gene expression can be engineered, we must first review the basic processes through which genes are natively expressed. The process through which genetic information is enacted upon in biological systems is often referred to as the “Central Dogma" of molecular biology. Coined in 1958 by Francis Crick[24], the simplified version of Central Dogma states that information flows in biological organisms from DNA through transcription into RNA, and from RNA through translation into proteins. While proteins form the majority of functional actors in biological systems, specialized RNA including transfer RNA (tRNA), ribosomal RNA (rRNA), small interfering RNA (siRNA), and antisense binding RNA (asRNA) are also crucial end products in the process of gene expression. Transcription of these RNAs and translation of messenger RNA (mRNA) into protein is the fundamental way in which genotypes give rise to phenotypes. Here I summarize how each of these processes works in the context of prokaryotes, focusing explicitly on gene expression in *Escherichia coli*. This sub-section is largely based on information from Chapters 29-32 of Voet and Voet’s 4th edition Biochemistry textbook[25].

2.1.1 Bacterial transcription

The first step of gene expression is transcription, in which DNA sequences are read by the cells’ transcriptional machinery and converted into RNAs. In *E. coli*, this machinery involves the RNA polymerase (RNAP) holoenzyme consisting of four protein subunits (α, β, β’, and the small, nonessential subunit ω) comprising a structural and

functional core, and a variable σ factor for which there exist various forms that each encourage binding to specific promoter motifs. The primary σ-factor is σ^{70} which ensures binding of the RNAP holoenzyme to essential genes, but six alternative forms exist. One of the most important alternative σ-factors is σ^{38} encoded by the *rpoS* gene, which encourages binding to genes related to stress response and will be important in future Chapters. Promoter elements generally consist of a -10, -35, and +1 sequence, indicating the DNA positions in relation to the first nucleotide transcribed from DNA into RNA (+1). RNAP binds to elements in the -10 and -35 DNA sequence of the promoter upstream of the +1 site, where transcription can begin.

The binding of RNAP is highly dependent on whether these promoter elements are free, which in turn is largely influenced by the binding of other proteins to DNA called transcription factors. Transcription factors interfere with the binding or processing of RNAP, or by altering the structure of DNA to encourage RNAP to promoter sites. Transcription factors are numerous, and enable bacteria to rapidly respond to their surroundings in the case of sudden environmental shock. As such, many transcription factors will be identified in the following chapters as promising candidates for targeting with gene expression engineering techniques to control bacterial phenotypes.

The first step is the formation of a transcription bubble by denaturing double-stranded DNA into accessible strands of single-stranded DNA. The strand which runs 5’ to 3’, called the sense or coding strand, is not directly transcribed. Rather, the strand running 3’ to 5’, called the antisense or template strand, is used by RNAP to seek out and attach complimentary nucleotides for performing transcription. The first nucleotide associated is the triphosphate RNA equivalent of the +1 sense nucleotide. RNAP then

elongates this nucleotide into a chain of complementary RNAs from 5’ to 3’. RNAP continues along the template strand making a complementary RNA until it reaches an intrinsic terminator sequence (generally long stretches of A’s and T’s), or a pause site where rho factor can end transcription.

2.1.2 Bacterial translation

RNAs freed from the transcription complex can then go on to perform their intended functions. For some genes such as tRNAs, this is the final stage of gene expression. For mRNAs however, they must now be converted into proteins through translation. This involves three stages performed by a ribosome: initiation, elongation, and termination.

In translation imitation, the 30S and 50S subunits of the ribosome are assembled on the target mRNA with the help of three initiation factors (IF1, IF2, and IF3) and N-formylmethionine (fMet). Translation in *E. coli* almost always starts at an AUG or GUG codon, with a short sequence ~3 codons upstream, called the ribosomal binding site, that exhibits complementarity to the 16s rRNA component of the 30S subunit. IF1 and IF3 bind to spent ribosomes at the 30S subunit to temporarily prevent 50S binding while the 30S subunit searches for a new ribosomal binding site. Then, IF2 and fMet-associated tRNA bind to the 30S, and work in tandem with IF3 to slot fMet into the position of the start codon. This causes IF1 and IF3 to disassociate, allowing the 50S subunit to join and form the ribosome. IF2 is disassociated, and translation elongation begins.

In elongation, tRNAs continue to bring the amino acid matching the codon for the next sequence in the growing polypeptide chain. This involves decoding the next codon,

transpeptidation of a new peptide bond, and translocation of the mRNA further along the mRNA. This process continues until a stop codon is reached, wherein a tRNA brings a release factor to cause disassociation of the mRNA/ ribosome/ polypeptide molecules. Thus, a new protein is formed, and the process of gene expression is completed.

In *E. coli*, and indeed most prokaryotes, transcription and translation are often coupled with one another. In eukaryotes, transcription of DNA into RNA occurs in the nucleus, where the RNAs are generally capped before export into the cytosol and translation at the endoplasmic reticulum. No such similar structure occurs in prokaryotes, where ribosomes are more free-floating.

Previous approaches to controlling gene expression have relied primarily upon altering the DNA sequence of particular gene's promoter and terminator sequences to alter the efficacy of transcription (and occasionally translation)[26–28]. However, direct modifications to promoters or ribosomal binding sites have proven time-consuming and difficult to rationally engineer[28]. Transcription activator-like effectors (TALEs) or zinc finger proteins (ZFPs) are highly effective, but notoriously difficult to engineer as new proteins must be created for every target sequence[29]. While gene expression can be increased by over-expressing the desired gene on a plasmid, this requires delivery of an entire plasmid circuit that can be difficult to express and maintain in non-traditional microorganisms[30]. In the next two sections, I describe two promising approaches to manipulate gene expression that overcomes these aforementioned limitations.

2.2 Regulating gene expression with CRISPR

2.2.1 A summary of the history of CRISPR

A revolution has swept through biology based on the discovery of clustered, regularly interspaced, short palindromic repeat (CRISPR) systems. These systems entail a family of nucleases and related cellular machinery that exist in prokaryotes. So far, at least 50% of bacterial genomes and 90% of archaeal genomes have been identified to contain CRISPR systems in some form[31]. The most basic summary of CRISPR systems is that they act as the immune systems of prokaryotes, providing a memory of previous viral infections and the machinery to cure the cell of infection upon subsequent exposures.

Like many biological systems, the discovery of CRISPR's existence significantly pre-dates the discovery of its function or importance. The first scientific publications recognizing the unique phenomenon of multiple tandem repeats in prokaryotes traces back to the ~1990s and two papers in particular. The first was in 1987 with Yoshizumi Ishino's investigation of the *iap* gene in *E. coli*[32], and the second was in 1993 with Francisco Mojica discovery of "a very peculiar DNA landmark" consisting of tandem repeat sequences in the archaea *Haloferax mediterranei*[33]. At the time, neither of these researchers recognized the implications of their discovery. By the early 2000's, it was becoming increasingly clear that repeat sequences were a prolific phenomenon across prokaryotic genomes. This was largely thanks to continued work by Mojica, who proposed the acronym CRISPR to summarize what he was seeing throughout prokaryotes[34,35].

The first major breakthrough in our understanding of the nature of CRISPR came in 2005 when Mojica reported that these repeat elements derived from foreign genetic elements[34,36]. This strongly suggested that CRISPR systems could be involved in immune

systems, and it was only a short matter of time until Rodolphe Barrangou proved that CRISPR provides resistance to bacteriophages[37]. This catalyzed a flurry of work to better understand CRISPR systems.

CRISPR's next major development was in 2012 when a collaboration between Emmanuelle Charpentier and Jennifer Doudna demonstrated that CRISPR-Cas could be co-opted for highly-specific in vivo genome editing[38]. This paper showed that the two RNA components directing CRISPR associated protein, the sequence-specific CRISPR RNAs (crRNAs) and scaffolding trans-activating crRNAs (tracrRNAs), could be synthetically combined into one functional chimeric RNA called a single guide RNA (sgRNA). This allowed the researchers to rationally engineer CRISPR to target virtually any genomic loci in a highly specific manner, exploiting the prokaryotic immune system for genetic engineering purposes. Since then, a majority of the "CRISPR craze" has focused on utilizing this breakthrough to alter DNA sequences in a vast array of organisms[39–43].

Two major CRISPR advances stemming from this original paper both came in 2013. In these papers, it was shown that deactivated versions of CRISPR-Cas with abolished nuclease activity (dCas9) could be utilized to inhibit[44] or activate[45] gene expression of virtually any gene. These discoveries were critical in extrapolating the use of CRISPR to rationally engineer gene expression, and form the basis of the CRISPR work utilized in this thesis. Also of importance is the very recent discovery in 2016 of CRISPR systems that target RNA instead of DNA sequences[46–48]. The following sections will further explore the structure of CRISPR systems, and their application to manipulate gene expression.

2.2.2 Structure of CRISPR systems

CRISPR systems are generally separated into two major categories: class I and class II, which utilize multiple or single CRISPR-associated enzymes (Cas) respectively to enact nucleic acid cleavage. These classes are further subdivided into subsections based upon the family of Cas enzymes, including types I, III, and IV in class I and types II, V, and VI in class II[49]. Attention has primarily been focused on the class II systems as single effectors are generally easier to implement. The hallmark CRISPR system is Cas9, a class II type II enzyme originally derived from *Streptococcus pyogenes*[50,51]. Cas9 is perhaps the most prolifically used CRISPR system in genome and gene expression engineering and was the system used by Doudna and Charpentier in their seminal CRISPR paper[38]. Another important class II enzyme is the type V Cpf1 from *Acidaminococcus sp. BV3L6*, which is a smaller enzyme that is rapidly replacing Cas9 in many engineering applications[52]. The final category in class II consists of type VI enzymes, generally called Cas13, that are distinguished by their ability to cut RNA in place of DNA[48]. Importantly, *E. coli* expresses only the class I type I system based on the Cascade complex of five Cas proteins, allowing for co-expression of the class II systems in *E. coli* chassis without interference of or from native CRISPR systems.

The defining feature of all CRISPR systems is the existence of a CRISPR locus harboring short repetitive elements separating unique DNA sequences. These unique sequences, referred to as spacers, are ~21-72 nucleotide DNA elements matching exogenous infections such as bacteriophages that the prokaryote has integrated into its own genome[53]. These spacers provide the “memories” of previous infections. The source DNA sequences for spacer elements are referred to as protospacers. These protospacer

elements generally exist next to unique 3-7 nucleotide sequences referred to as the protospacer-adjacent motif (PAM)[53]. These PAM sites are important for recognition by CRISPR complexes; even if the guide RNA perfectly matches a nucleic acid sequence, Cas enzymes from classes where PAM sites are involved in protospacer acquisition will not bind to and cleave the target if the PAM sequence is absent.

The second major component of CRISPR systems are the Cas enzymes that effect cleavage of exogenous genetic elements. The joining of these Cas enzymes with guide RNAs allows creates a protein-RNA complex that seeks out specific nucleic acid sequences for subsequent cleavage. Many CRISPR systems require additional processing enzymes to generate usable guide RNAs, such as the class II type II system which requires a separate enzyme from Cas9 to process crRNA and tracrRNA into the final guide RNA that complexes with Cas9.

The general process through which prokaryotes express their CRISPR systems begins with transcription of the crRNAs and any associated RNAs such as tracrRNA, processing of the RNA through helper proteins or the final Cas enzymes themselves, and formation of the Cas nuclease with the processed guide RNAs into an effector complex[54]. The Cas-RNA complex then searches out for target sequences by scanning DNA for PAM sites. Once found, the Cas complex begins to unwind the target DNA, allowing for complexing of the target single-stranded DNA with the complementary guide RNA. Important in this process is the “seed sequence”, a subsection of the guide RNA where the formation of the DNA-RNA bounds is initiated[31]. Depending on the local thermodynamics of the DNA sequence, one mismatch in this seed sequence may be enough to completely abolish Cas binding. However, it has been demonstrated that Cas

enzymes can tolerate up to three mismatches within the seed sequence before losing the ability to bind to its target, albeit with a noted reduction in binding efficiency as mismatches accumulate[55]. If the sequence is a match, Cas enzyme(s) initiate cleavage of the target DNA. A schematic of this process is presented for Cas9 in Figure 1, as Cas9 is the poster Cas enzyme. It should be noted that Cas9 utilizes an “NGG” PAM sequence, where “N” can be any nucleotide. There is also still some slight ability of Cas9 to bind to DNA sequences with a “NAG” PAM site, although the binding is significantly weaker.

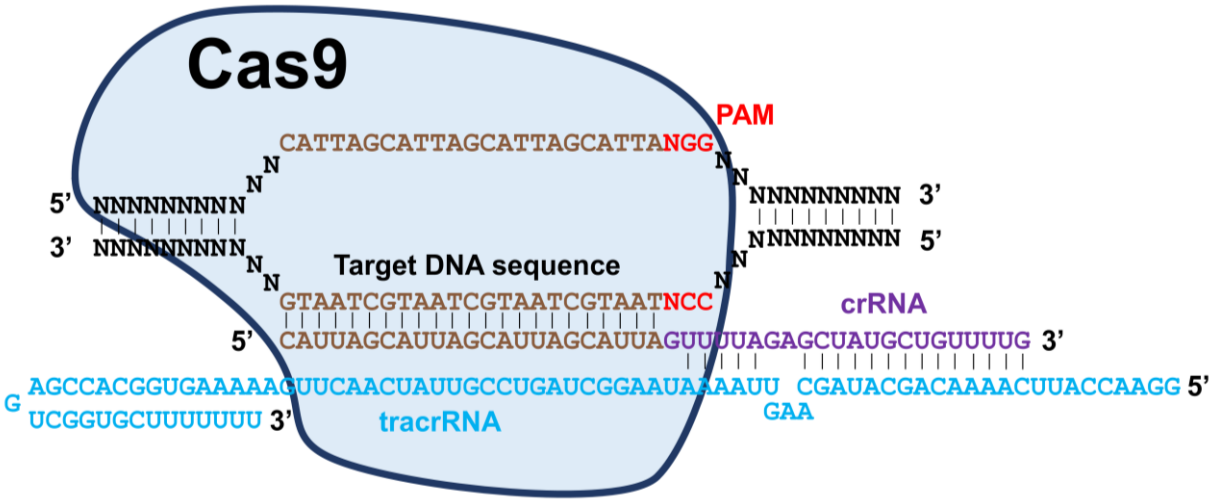


Figure 2.1 Cas9 association to target DNA sequence

An outline of how Cas9 associates with a target DNA. For this system, the PAM motif is always located on the 3’ end of the opposite strand to which the crRNA binds. In the native system, a secondary tracrRNA is required for complexing of the crRNA to the Cas9. A chimera of these two RNAs into a single guide RNA (sgRNA) was accomplished by Jinek et. al., and is the preferred guide for engineering applications[38].

2.2.3 Using dCas9 and dCas9-ω to tune mRNA production

One of the implications of the CRISPR revolution is the development of novel techniques for controlling gene expression in virtually all classes of organisms. Although these systems have evolved for the purpose of degrading targeted nucleic acids, clever

scientists have genetically engineered these natural systems to enable controlled modification of gene expression without impacting DNA sequence or structure.

The seminal study demonstrating the application of CRISPR to control gene expression was published in 2013 by Lei Qi and Wendell Lim[44]. In this study, the authors utilized a catalytically dead Cas9 variant (dCas9) as their effector protein. This dCas9 variant exhibits two point mutations in the RuvC1 and HNH domains of Cas9 that are responsible for performing the nuclease activity of the enzyme. As such, the protein is unable to cut DNA. Despite this, dCas9 still retains its ability to search out and bind to DNA elements when provided with a sgRNA chimera. In doing so, dCas9 is able to act as a synthetic transcription factor that can bind to virtually any location in a genome.

Indeed, the authors demonstrated that dCas9 could be used to knockdown initiation and elongation of transcription, thereby reducing the overall expression of a targeted gene. Targeting dCas9 to the non-template strand at the start of gene's open reading frame (ORF) resulted in roughly 300-fold repression, while targeting the template strand near the -35 sequence of a gene reduced expression ~100-fold. While targeting other regions also resulted in significant gene repression, these areas were shown to be the most effective places to locate to enact maximum gene repression. The most likely explanation for these impacts on gene expression is a roadblock mechanism for targeting the ORF and occlusion for targeting the promoter. In the first case, RNAP runs into dCas9 while progressing across the DNA and is unable to continue, and therefore falls off the DNA terminating transcription. In the latter case, RNAP is unable to even bind to the DNA sequence due to the presence of dCas9 in its binding socket[56]. This study also showed that by optimizing the dCas9 sequence for expression in human cells and tagging a

nuclear localization sequence to the C-terminus, dCas9 could be used to repress gene expression in human cells. These results were optimized in a study also published in 2013, in which Qi published another paper with Luke Gilbert demonstrating that fusing dCas9 to the Kruppel-associated box (KRAB) domain of Kox1 significantly enhanced repression of gene expression in HEK293 cells[57]. In this manner, Qi and others established the technique now known as CRISPR interference, often referred to as CRISPRi. It should be noted that CRISPR interference originally refered to degradation of plasmid DNA[58], but this terminology has fallen out of favor in literature.

Just a few months after Qi's first study was published, another study done by David Bikard and Luciano Marraffini demonstrated that dCas9 could be utilized not only to repress gene expression, but to also activate gene expression[45]. This work first conveyed the same results as Qi's paper but showed a more consistent level or repression of gene expression when dCas9 was targeted anywhere between the -35 promoter sequence up to ~50 nucleotides inside the ORF, regardless of which strand was targeted. The key novelty of this work came in its demonstration that dCas9 could be tagged with the ω-subunit of RNAP and that this dCas9-ω variant could subsequently be used for increasing gene expression. As the ω-subunit is involved in the formation of the complete RNAP holoenzyme, attaching it to dCas9 helps to recruit RNAP to particular genomic locations by stabilizing its binding to the promoter sequence. The authors demonstrated that targeting dCas9-ω upstream to the promoter sequence caused an increase in gene expression, finding that a range of 80-100 nucleotide distance between the PAM and the transcription start site resulted in optimal gene activation. However, this level of gene expression was significantly lower than the level of gene repression potential that dCas9

exhibits. In the best circumstance, the authors achieved a 23-fold increase in gene expression. Furthermore, the authors performed these experiments in a strain of *E. coli* in which the ω gene (*rpoZ*) was deleted, likely to increase the complementarity between dCas9-ω and RNAP which can form even if the ω subunit is missing. Nevertheless, these results served to revolutionize the field of CRISPR control of gene expression by providing the first tool for activating gene expression. Subsequent studies have demonstrated that CRISPR activation can also be achieved in human cells by fusing other native transcriptional activators to dCas9, such as VP64[59]. In keeping with the connection to CRISPRi, the process of activating gene expression with CRISPR-transcriptional activators is now referred to as CRISPR activation or CRISPRa. An example of how dCas9 and dCas9-ω are used to regulate gene expression is shown in Figure 2.

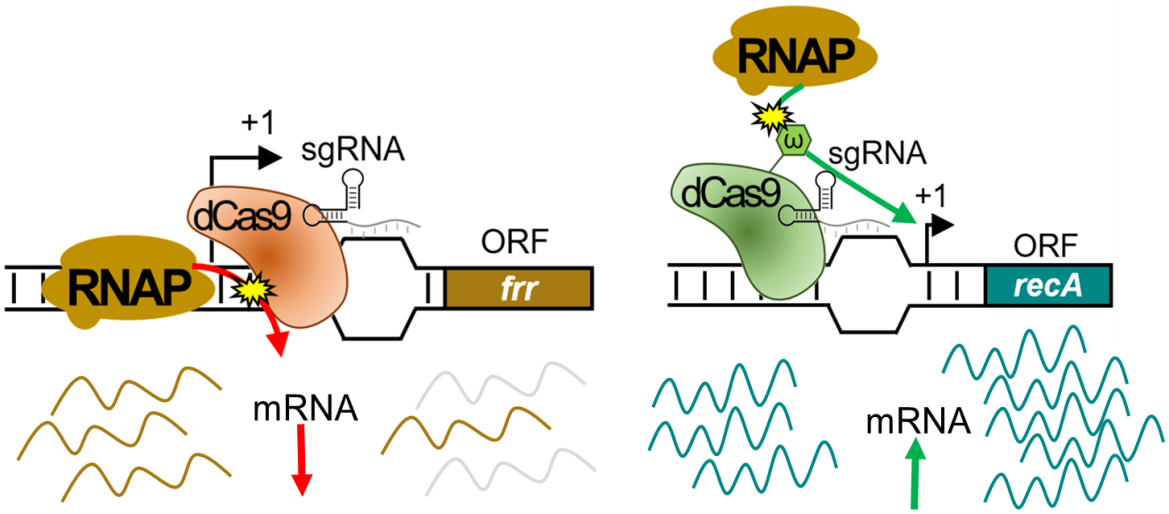


Figure 2.2 CRISPR regulation of gene expression

Schematic of deactivated CRISPR associated protein 9 (dCas9) and its use for decreasing gene expression (left) or increasing gene expression (right) upon fusion to the ω-subunit of RNA polymerase.

Since 2013, CRISPRi and CRISPRa have rapidly proliferated in use, although not to as great an extent as CRISPR for DNA modification. In human cells, CRISPRi and

CRISPRa have systematically explored a genome-scale perturbation of gene expression to identify important genes for cell phenotypes such as survival and differentiation[60]. CRISPR perturbation has also been utilized to impart complex transcriptional control using AND, OR, NAND, and NOR logic gates[61]. CRISPRi has also been demonstrated to successfully knockdown expression of specific genes in the mouse hippocampus, showing its potential for use in altering brain functions without direct genome modifications[62].

In microorganisms, it has been shown that multiple guide RNAs can be included in an array in order to target dCas9 to multiple genes at once[63] through a multiplexed fashion[64]. Endogenous systems have been co-opted for repurposing a bacteria to repress expression of its own genome[65]. In terms of applications, CRISPRi has had some success in improving antimicrobial treatment, where it has been used to knockdown *luxS* expression and thereby reduce biofilm formation[66]. But perhaps the most sought-after application of CRISPR perturbations in microorganisms is in metabolic engineering of native of synthetic pathways for optimization of end product production. For example, CRISPRi has been applied to optimize *Saccharomyces cerevisiae* consumption of glycerol and production of 3-dehydroshikimate[67], to improving the biosynthesis of amino acids related to n-butanol synthesis[68] in *Klebsiella pneumonia*, and to enhancing titers of 1,4-butanediol in *E. coli*[69]. In Chapters 4-7, I will explore novel ways that CRISPRi and CRISPRa have been applied in our lab in relations to both antimicrobial treatment and generation of bioproducts.

2.2.4 RNA targeting with C2c2/Cas13

A very nascent discovery in the field of CRISPR is the enzyme Cas13. This enzyme was originally identified through a thorough phylogenic analysis performed in 2015 by Shmakov and Koonin[70]. In this study, the authors investigated the genomes of prokaryotes closely related to those expressing the known class 2 systems of Cas9 and Cpf1. They discovered three new previously uncharacterized class 2 systems, the most promising of which was the system known as class 2 candidate 2 or C2c2. They saw that the CRISPR systems in this class exhibited two predicted HEPN RNase domains, raising the possibility that these CRISPR systems might function on the RNA level in lieu of the commonly utilized DNA level by other CRISPR systems.

Indeed, a study just one year later done in collaboration with these authors and Feng Zhang showed that C2c2 CRISPR systems indeed targeted RNA[48]. Discovered in *Leptotrichia shahii*, the authors demonstrated that this enzyme was able to degrade single-stranded RNA when provided with a complimentary guide RNA. This was a breakthrough in CRISPR technology, as a single-effector CRISPR protein for targeting RNA had been undiscovered up to this point. The authors demonstrated that like most other CRISPR systems, C2c2 relies on the existence of a motif adjacent to the protospacer in order to enact effective binding to RNA. However, to distinguish this system from other class 2 systems, the authors refer to this as a 3' protospacer-flanking site or PFS instead of a protospacer adjacent motif or PAM. They showed that this PFS site is much more lenient, requiring only the avoidance of a G nucleotide in the site immediately adjacent to the protospacer. A slight preference for C's was also found in positions 1, 2, and 5 of the PFS. The authors also demonstrated that the C2c2 system is

both a single effector (one enzyme) system, as well as a single CRISPR RNA (crRNA) system requiring no additional tracrRNA. Most intriguingly, the authors demonstrated a unique process through which C2c2 degrades its target RNA. An apparent “collateral” effect was observed in which not only was the target RNA degraded, but also other RNAs present in the mixture. Notably, this was not the effect of random nonspecific targeting of the C2c2-crRNA complex, as no collateral effect was observed in the presence of a non-targeting crRNA control. It is likely that the CRISPR complex is activated upon binding to its target RNA, changing confirmation into an enzymatic “bomb” that destroys all RNA around its target location.

Since 2016, C2c2 has been renamed in literature as Cas13a. Continuing work into Cas13a functionality has demonstrated that Cas13a is indeed able to process its own crRNA into mature, fully functional crRNAs[71]. The crystal structure of Cas13a bound to its target crRNA has also been resolved, demonstrating that the two HEPN sites in Cas13a are present for degradation of target RNA, while a separate site in the Helical-1 domain is responsible for processing its own crRNA[71]. Additionally, it has been shown that processing of this crRNA is not essential for proper Cas13a degradation of target RNA[72]. A number of applications have been demonstrated for Cas13a, including RNA editing of mammalian transcripts[47], detection of specific RNAs using catalytically dead Cas13a (dCas13a) in a highly sensitive fashion[73], and degradation of RNA viruses in plants[74].

Only two years have passed since the discovery of this novel CRISPR system, and yet the potential applications of Cas13a are vast. In Chapters 8 and 9, I will explore two novel applications of Cas13a that our lab has begun working towards.

2.2.5 Delivery systems for CRISPR

A significant barrier to applying CRISPR systems to manipulate gene expression is biological barriers themselves. Specifically, cell membranes are notoriously difficult for large proteins to traverse, and those that do often lose their tertiary structure and are therefore made ineffective once localized inside the cell[75]. The most simplistic approach to deliver CRISPR systems is through electroporation or microinjection of CRISPR DNA into the target cell[50]. However, these approaches are impractical outside of laboratory conditions. As such, specialized techniques must be employed in order to deliver Cas enzymes and their corresponding guide RNAs to their target cells.

One approach is to use one bacteria to deliver DNA sequences encoding the CRISPR system to another bacteria via conjugation. Conjugation is the process through which one bacteria forms a membrane bridge with another bacteria, through which plasmids or other transposable genetic elements can be shared between bacteria[76]. Work in our lab has demonstrated that such an approach can be used to deliver a novel antimicrobial treatment (Collins *et. al.*, submitted). An iGEM team showed that a CRISPRi system could be delivered from one *E. coli* strain to another and effectively knockdown gene expression in the target strain[77].

A similar approach to that presented above is to use a bacteriophage for delivery of CRISPR systems. Bacteriophage are virus particles that infect specific strains of bacteria and have long been considered for the potential to deliver DNA and other nucleic acids to bacteria, especially in the former Soviet Bloc[78,79]. In 2014, Bikard and Marraffini demonstrated that Cas9 could be delivered alongside a sgRNA on a phagemid, a plasmid

that replicates and packages inside bacteriophages, into *Staphylococcus aureus* in a mouse skin colonization model. They showed that this approach effectively killed virulent bacteria in a sequence-specific fashion, leaving avirulent bacteria unperturbed[80]. Another study demonstrated that bacteriophage delivery of Cas could also be used to selectively kill antibiotic-resistant bacteria[81]. A great body of work has been generated in using a complementary approach to deliver CRISPR systems in eukaryotic systems based upon adenoviral vectors, non-integrating viral carriers of genetic information[50].

The above approaches share one significant drawback: they rely upon the target cell's transcriptional and translational machinery to express the CRISPR systems in the first place. This raises the potential for errors in transcription, translation, or replication of the plasmids that abolish the system's efficacy. As such, approaches for direct delivery of fully-assembled CRISPR systems are of substantial interest. One way this can be accomplished is through conjugation of CRISPR machinery to cell penetrating peptides or CPPs. CPPs are a class of short peptide motifs that have the ability to drastically increase the transportability of a number of molecules, including proteins and RNAs[82]. Depending on the specific sequence, CPPs function by promoting either endocytosis into eukaryotic cells, or promoting direct translocation of the molecule they are conjugated to across the membrane[82]. This approach was employed to deliver Cas9 and sgRNAs simultaneously into human cell lines while maintaining CRISPR functionality[83]. This CPP approach is attracting significant interest, as it could likely be replicated to deliver CRISPR machinery for the manipulation of gene expression[84].

Another promising approach for delivery of CRISPR systems is to use nanoparticles loaded with Cas protein and sgRNAs as a vehicle to traverse the

membrane. It has been demonstrated that rolling circle amplification can create DNA-based nanoparticles that self-assemble into a spherical structure with partial complementarity to the sgRNA sequence to encourage Cas-sgRNA complex binding. These molecules were shown to successfully deliver a Cas9-sgRNA into mice[85].

2.3 Manipulating gene expression with peptide nucleic acids

2.3.1 PNA chemistry and applications

Peptide nucleic acids or PNAs were invented in 1991 by Peter Nielsen, Michael Egholm, and Ole Buchardt[86] as a way of demonstrating our ability to create synthetic nucleic acid molecules. In this study, the authors replaced the ribose-phosphate backbone of DNA with a pseudo-peptide polyamide backbone while maintaining the ACGT nucleosides, mimicking the structure of DNA. They showed that this single-stranded DNA mimic was able to displace double-stranded DNA and bind with "extraordinarily high stability" to its complementary sequence. PNA acts as a remarkably effective DNA mimic because of its constrained flexibility that forces the molecule into a helical structure preferred by nucleic acids[87]. Indeed, PNAs exhibit more affinity for RNA and DNA than natural RNA-RNA and DNA-DNA interactions, as they have a neutral backbone instead of the charged backbones of nucleic acids[88].

Interestingly, it has been suggested that PNA molecules may have been the first genetic molecule ever to have evolved and has been demonstrated to be produced under pre-biotic conditions through natural processes such as electronic discharge[89]. It has even been demonstrated that PNA-PNA molecules can form double-stranded helical duplexes in a similar fashion as DNA-DNA[90]. However, PNAs are not known to be

produced in any biological systems. As such, there are no known enzymes that are able to cleave PNAs, making them a highly attractive molecule for many applications. Indeed, PNAs have been shown to be highly stable and resistant to a wide range enzymatic degradation in *in vivo* contexts[88].

Perhaps the most sought-after application for PNA is in antisense RNA therapy[91]. As both RNA and PNA are single-stranded molecules, those with complementarity to one another will bind readily. Upon binding to RNA, PNA has been shown to block translation through occlusion of the ribosome or by blocking the procession of the ribosome across mRNA[92]. Furthermore, the length of PNA required to accomplish this translation knockdown is relatively short, and generally in the range of 10-15 nucleotides long. This capability of PNA has spurred significant interest in its use to block expression from RNA *in vivo*, which allows for the possibility of using PNA as a sequence-specific therapy based on manipulating gene expression. A number of studies have demonstrated the use of PNA molecules as a therapy against a variety of different bacterial strains[93–95]. It has also been shown that PNA is not readily removed from bacteria through natural processes such as efflux pumps[96]. These PNAs have even been used to resensitize antibiotic-resistant bacteria to antibiotic therapy[97].

Additionally, PNA has been employed to track RNA and other molecules *in vivo*. For instance, PNAs has been coupled with fluorescence *in situ* hybridization (FISH) to track biofilms in a mixed population[98] or for identifying *Staphylococcus aureus* exhibiting coagulase-negative phenotypes[99]. PNAs have also been coupled with single-stranded DNA to identify multivalent interactions from the simultaneous binding of multiple ligands to a molecule with multiple receptors[100].

2.3.2 Delivery of PNAs

Despite the potential of PNA for sequence-specific manipulation of gene expression, there are two major barriers to its application. The first is its relatively high cost of synthesis, as there are no biological systems that can be exploited to rapidly generate them. The second is the poor ability of PNA to traverse biological membranes. This is due to the high mass of these molecules and the lack of any significant charge, causing significant steric hinderance[101]. While the cost issue is still a significant obstacle, a number of strides have been made in improving the ability to deliver PNA molecules to various biological systems.

Perhaps the most prominent way in which PNAs are delivered is through is through conjugation to CPPs. This was demonstrated in 2001 when Lima Good and Peter Nielsen attached the KFFKFFKFFK peptide motif to a PNA targeting both rRNA and *acp* mRNA and showed that this motif greatly improved transport of PNAs into *E. coli*, thereby increasing the efficacy of translation interference[102]. Since this demonstration, the focus on CPP-mediated PNA delivery has been primarily given to eukaryotic systems. For example, the addition of a few lysine residues has been shown to significantly increase endocytosis of PNA molecules into eukaryotic cells[103]. The same work previously mentioned also demonstrated that PNAs can be complexed with DNA oligonucleotides and a cationic lipid to mediate endocytosis[103]. Once inside mammalian cells, PNA still faces the potential obstacle of transport into the nucleus to bind to RNAs before they are exported into the cytoplasm. This barrier has been addressed through previous work showing that PNA activity can be improved by conjugation to nuclear localization signal (NLS) such as PKKRKV[101,104].

Other novel approaches are being investigated for delivering PNA to both prokaryotic and eukaryotic systems. For instance, electroporation has also been coupled with PNA-peptide conjugates to improve PNA delivery to intracellular infections[105]. The use of nanoparticle molecules has also been explored for delivery of PNAs. Such nanoparticle spheres are coated with a thin layer of gold and used as a platform for PNA binding and delivery[106].

Previous work in our lab has shown that delivery of PNAs to bacteria is readily accomplished through conjugation to the canonical KFFKFFKFFK sequence demonstrated by Good and Nielsen[97,107]. Indeed, a number of other studies have utilized this sequence to deliver antisense PNAs to a wide range of bacteria effectively[95,96]. PNAs conjugated to a KFF motif have been shown to effectively kill bacteria at lower concentrations than other CPP motifs, suggesting that this motif is one of the most effective sequences for delivery of PNA into bacteria[108].

Chapter 3: Elucidating bacterial gene expression response to antibiotic and biofuel stress conditions

Reprinted in part with permission from Erickson, K. E., Otoupal, P. B., and Chatterjee, A. (2015) Gene Expression Variability Underlies Adaptive Resistance in Phenotypically Heterogeneous Bacterial Populations. ACS Infectious Diseases, pg. 555-567. DOI: 10.1021/acsinfecdis.5b00095. *© 2015 American Chemical Society. Accessible at https://pubs.acs.org/doi/10.1021/acsinfecdis.5b00095*

Reprinted in part with permission from Erickson, K. E., Otoupal, P. B., and Chatterjee, A. (2017) Transcriptome-Level Signatures in Gene Expression and Gene Expression Variability during Bacterial Adaptive Evolution. mSphere (vol 2), pg. 555-567. e00009-17. © 2015 American Society for Microbiology. Published under the Creative Commons CC BY license. Accessible at http://msphere.asm.org/content/2/1/e00009-17

3.1 Abstract

In nature, microorganisms rarely have the luxury of existing in an ideal environment. Instead, the environments in which they live are prone to frequent and sudden fluctuations. This necessitates that microorganisms establish biological strategies to provide the phenotypic flexibility needed to survive such environmental uncertainty. One such strategy that microorganisms have evolved is a greater level of stochasticity in gene expression relative to higher order systems. This allows subsets of a population to be constantly primed for survival in case the environment shifts, and also allows rapid adaptation to persistent exposure in stressful environments. In this Chapter, I summarize two studies in which we explored the natural processes by which bacteria gene expression is shaped by exposure to stressful environmental conditions, including antibiotics and toxic biofuels. The first study demonstrated the presence of extensive

inter- and intra-population phenotypic heterogeneity across adapted populations, and identified through clustering analysis that inter-population gene expression variability in adapted populations was distinct from that of unadapted populations. Sequencing select genes revealed that the observed gene expression trends are not necessarily attributable to genetic changes. In the second study, we investigated the transcriptome profiles of bacteria upon adaptation under biofuel and antibiotic stress conditions to locate potential target genes via conventional gene expression analysis and by examining differential gene expression variability. In both these studies, we validate the biological relevance of the observed gene expression changes by synthetically perturbing gene expression using CRISPRi and CRISPRa. Together, these works provide evidence for the importance of gene expression variability in responding to environmental stressors, and establishes a set of promising gene candidates to be manipulated in future Chapters. This Chapter is adapted in large part from work led by Dr. Keesha Erickson[16,17,109].

3.2 Introduction

Nature is rarely static. Life is constantly exposed to diverse and stressful environments that necessitate rapid response strategies for survival. In microorganisms, this has led to the evolution of bet-hedging, a process in which microbes exhibit altered gene expression to maximize fitness under new conditions[110–112]. Heterogeneity at the gene expression level is credited for making this possible, allowing for resilient subpopulations that survive conditions normally lethal to the population at large[113–115]. Bacteria, in particular, are noted for their ability to rapidly adapt to myriad environmental pressures[14], by taking advantage of this non-genetic response[116]. This phenomenon has been observed for many conditions across a variety of bacterial species[117–121]. Intra-population heterogeneity allows adapting populations to sample multiple states, without genetic alteration, to maximize the probability of survival.

Rationally tuning heterogeneity in gene expression could be a significant avenue by which to improve or diminish the ability of bacteria to survive sudden stress exposure. However, our knowledge of which genes are key players in this resilience phenomenon is currently significantly limited. The main challenge in expanding our knowledge of gene expression heterogeneity is the vast scale at which it occurs. And while much effort has been devoted to scrutinizing mutational trends during adaptation[111,122–125], only more recently have studies emerged considering non-genetic contributions to resistance[126,127]. Several studies indicate that stress response genes tend towards noisy gene expression characteristics[128–130]. Even in individual cells of *E. coli*, significant stochasticity in gene expression has been identified[131]. The potential influence of epigenetic modifications further obfuscates any attempt to try and parse this response[132]. Given these

considerations, distinguishing what noise is important for survival, and what noise is simply ancillary, is a daunting task.

In this Chapter, we explore bacterial population-level gene expression heterogeneity and its relationship with the adaptive response of these organisms to stress exposure. We specifically investigate how *E. coli* gene expression changes during exposure to two antibiotics (ampicillin and tetracycline), as well as two biofuels (n-butanol and n-hexane). We begin this by probing population-level gene expression patterns in specific stress response pathways including the *mar* regulon and the SOS response, as well as general stress response genes[16]. RT-qPCR results reveal significant inter-population heterogeneity in expression of these genes, as well as key conserved gene expression responses. We show that this gene expression variability in populations adapted to biofuel stresses is different than the gene expression variability in populations not exposed to stress. Using CRISPRi constructs, we show a correlation between the degree of gene expression variability and the ability of strains to adapt to stress exposure. We then employ a *de novo* approach to determine global transcriptional changes during stress exposure[17]. Adapted populations of duplicate *E. coli* K12 MG1655 samples were exposed to ampicillin, tetracycline, and n-butanol for 11-14 days, after which expression profiles were determined using whole-genome RNA-Seq. We identified genes who exhibited unique genetic signatures, including significantly lower or higher variability either within replicates or across populations. This pointed us towards interesting pathways involved in the natural response of bacteria to stress exposure. Manipulation of these pathways could serve as a way to deter survival during antibiotic exposure, or conversely, increase tolerance during biofuel production.

3.3 Results

3.3.1 Inter-population variability in gene expression during E. coli exposure to biofuels

We began by investigating how *E. coli* MG1655 populations exposed to n-hexane and n-butanol naturally responded to n-butanol and n-hexane at the gene expression level. While it has been shown previously that *E. coli* K12 derivatives are intrinsically resistant to n-hexane via AcrAB-TolC efflux[133], n-butanol is known to be toxic to *E. coli*, disturbing cellular respiration, metabolism, and transport, as well as activating the *mar* regulon, the oxidative stress response, the membrane stress response, and the heat shock response[134].

As *E. coli* readily survive exposure to high concentrations of n-hexane, we measured gene expression of three populations exposed to 10% v/v n-hexane for 30 hours. For n-butanol, we adapted three populations for 14 days, beginning at 1% v/v n-butanol. Populations were inoculated in both 1.0% and 2.0% v/v n-butanol at the beginning of each new day, with the highest concentration survived used to inoculate new cultures the next day. While some populations temporarily survived high concentrations, by the end of the experiment all three had only grown in 1.0% v/v n-butanol. These end cultures were used to determine gene expression profiles in n-butanol.

Figure 3.1 shows qPCR results of the gene expression profiles of 14 stress response genes during exposure to (A) n-butanol and (B) n-hexane. In n-butanol adapted cultures, *hfq* (P=0.02), *lexA* (P=0.04), and *rpoS* (P=0.06) were differentially under-expressed, while *recA* (P=0.01) was differentially over-expressed. Few genes were tightly expressed, while others were variable across the individually adapted populations. In n-

butanol cultures, *tolC*, *rpoS*, *hfq*, and *recA* gene expression was less variable across the adapted populations, while the most variable genes were *cyoA* and *soxS*. In contrast, none of the genes were significantly under-expressed in the n-hexane populations, though *rob* (*P*=0.08) and *soxS* (*P*=0.09) were over-expressed. *cyoA* expression had the lowest variability across the n-hexane populations, in striking contrast to the high expression variability in *cyoA* across populations adapted to ampicillin, tetracycline, or n-butanol. Other genes with lower variability in n-hexane included *rob, soxS,* and *recA*, while *lexA*, *marA*, and *dinB* had the largest inter-population variability.

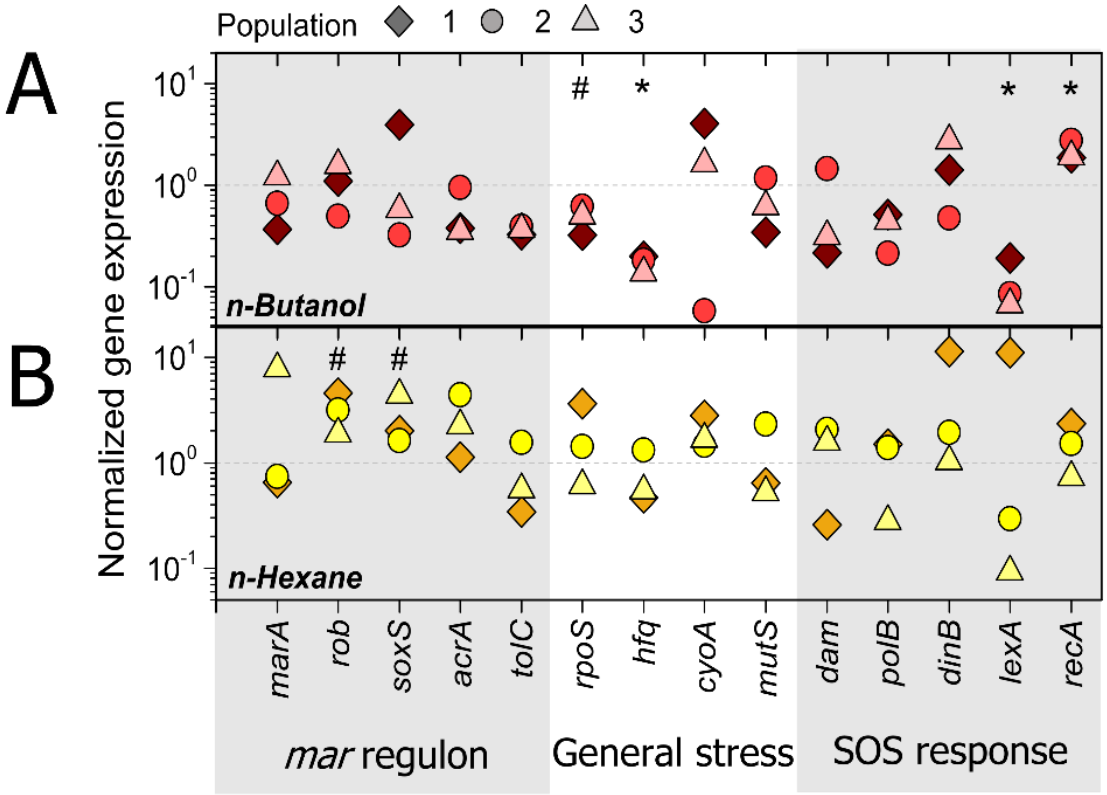


Figure 3.1 Gene expression response of *E. coli* exposed to n-butanol and n-hexane

Normalized expression of fourteen stress response genes is shown for n-butanol adapted populations (A) and populations grown with n-hexane (B). Normalized gene expression was calculated as described in the Methods. The horizontal line at 1 marks expression equal to wild-type. Significant differential expression is denoted with * for *P*<0.05 and # for *P*<0.10.

3.3.2 Shifts in gene expression variability during stress exposure.

We next sought to compare the gene expression profiles of populations before and after exposure to biofuel and antibiotic stress conditions. As n-hexane exposure was only performed over 30 hours, this population was considered non-adapted. qPCR for wildtype MG1655 not exposed to any stress condition was performed to provide a prime example of the variability in gene expression of unadapted *E. coli*. Additionally, three populations of MG1655 were exposed to either ampicillin or tetracycline for 11 to 14 days, with sequential increasing of antibiotic concentration at the beginning of each new day as was done for n-butanol.

We examined the inter-population variability across adapted populations by employing hierarchical clustering of the range of gene expression changes (Fig 3.2). Notably, the unadapted and adapted populations were divided into separate clusters using hierarchical clustering. Among the adapted populations, the tetracycline and ampicillin adapted populations were closest together.

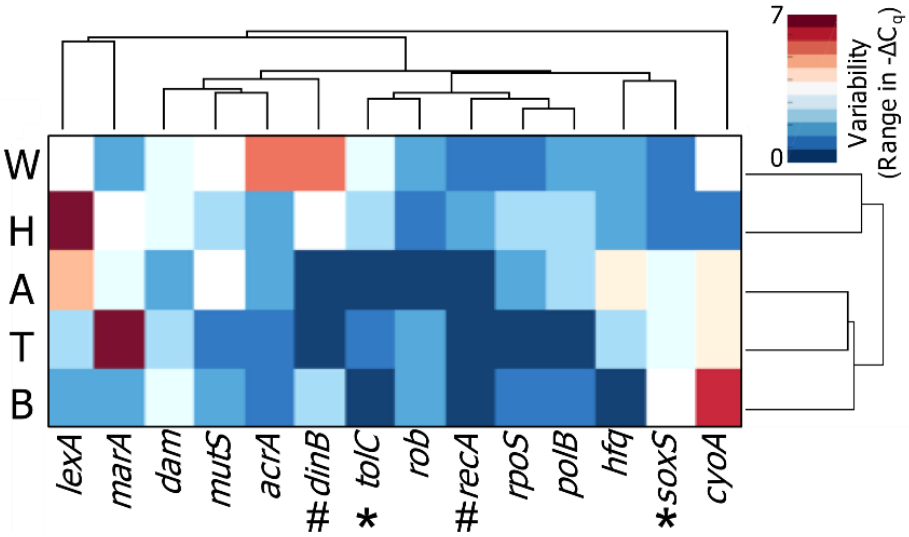


Figure 3.2 Hierarchical clustering of gene expression profiles before and after exposure to biofuels and antibiotics

Heatmap values indicate the inter-population range of gene expression ($-\Delta C_{q,avg}$) from each of the five growth conditions (shown on vertical axis): wild-type (W), adapted to ampicillin (A), tetracycline (T), or n-butanol (B), or grown with n-hexane (H). Hierarchical clustering is based on Euclidean distance. Significantly different expression variability between unadapted and adapted populations is indicated with * for $P<0.05$ and # for $P<0.10$.

Hierarchical clustering highlights patterns in gene expression variability that separate adapted and unadapted populations. Importantly, we find that in adapting populations while few genes exhibit increased inter-population gene expression variability, few other genes demonstrate decreased inter-population gene expression variability with respect to wild-type or unadapted hexane population. Thus, the overall expression variability pattern across the set of genes is fundamentally different for adapted versus unadapted populations. Interestingly, several genes were significantly differentially variable in adapted conditions versus unadapted conditions, perhaps signifying a stress-specific adaptive resistance mechanism.

Genes undergoing significant reductions in gene expression variability upon adaptation suggest that a specific, small range of expression levels is selected for even in divergent adapted populations under different stress conditions. One of the most variable genes in the wild-type samples was *dinB*, for which variability significantly decreased in adapted cultures ($P=0.07$). Similarly, for *tolC* and *recA,* the adapted populations displayed tighter expression than the unadapted populations ($P=0.02$ and $P=0.08$ respectively). In contrast, *soxS* expression variability was significantly higher across adapted populations than across wild-type and hexane treated strains ($P=0.01$), signifying that a high range of expression levels is tolerable for adaptation.

Clustering also allowed identification of groups of genes exhibiting similar trends in variability. For example, *lexA* and *marA* were grouped separately from the bulk of the genes, and were relatively more variable in n-hexane, tetracycline, and ampicillin adapted samples. Another noteworthy grouping contained *tolC*, *rob*, *recA*, *rpoS*, and *polB*, which were generally expressed with lower inter-population variability. Previous studies have observed that several of these low-variability genes play a direct role in survival to diverse antibiotics; for instance, *tolC* in presence of ciprofloxacin, sulfamethoxazole, metronidazole, and ampicillin[135], and *recA* in presence of ciprofloxacin, sulfamethoxazole, metronidazole norfloxacin, ampicillin, and kanamycin[135,136]. Additionally, in *E. coli*, *Pseudomonas aeruginosa*, and *Vibrio cholera*, inhibition of *rpoS* was shown to decrease the frequency of resistant colonies upon exposure to ampicillin[137]. Finally, identifying differences between the adapted populations from each condition may provide insight into stress-specific mechanisms of adaptive resistance. When we compare the tetracycline and ampicillin adapted populations, *marA* expression was relatively more variable across tetracycline adapting strains than ampicillin-adapting strains, while *lexA*, *polB*, and *mutS* were more variable across ampicillin cultures.

3.3.3 Gene expression variability impacts adaptation ability

To scrutinize the relationship between expression variability and adaptation, we applied the synthetic CRISPRi (Clustered Regularly Interspaced Short Palindromic Repeats interference) system[44] to manipulate expression of seven genes: *marA*, *acrA*, *tolC*, *dinB*, *soxS*, *recA*, and *mutS*. CRISPRi blocks transcription via interference from the deactivated RNA-guided DNA endonuclease, dCas9, which binds to DNA in a region

specified by the 20-nt sequence of a single guide RNA (sgRNA). The sgRNA sequences were designed to guide dCas9 to an NGG PAM sequence in the open reading frame for each target gene. Plasmids expressing the sgRNA and dCas9 protein were co-transformed into *E. coli* MG1655. A sgRNA plasmid targeting red-florescent protein (RFP), which is not present in *E. coli* MG1655, was also transformed as a control strain. The degree of repression from all constructs was validated with qPCR (Fig. 3.3). The CRISPRi strains were grown in media without or with 1 μg/mL of tetracycline for a period of three days to evaluate the impact of gene perturbation on adaptation.

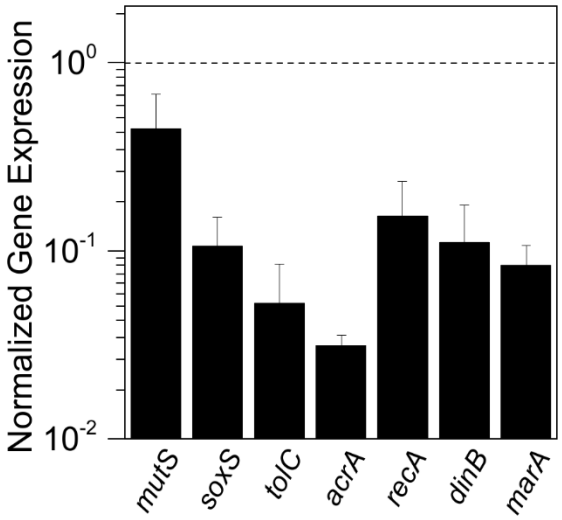


Figure 3.3 qPCR validation of CRISPRi strains

Normalized gene expression is shown for the target gene in each CRISPRi strain (x-axis). Expression is calculated relative to control strain *E. coli* MG1655. Error bars are the SD of n=3 independent biological replicates.

To quantify adaptation ability, we calculated an adaptation factor ($\alpha_{m,c}$), which describes the average change in growth rate upon gene perturbation relative to the control ($r_{m,c}$) and relative to the initial growth rate (on day 1) upon gene perturbation ($s_{m,c}$)

(Equations 3.1-3). The subscripts "*m*" and "*c*" denote the CRISPRi strain and adaptation condition (either tetracycline or no toxin), respectively. $r_{m,c}$ was calculated as the sum of relative change between the growth rate of the specific CRISPRi strain (μ_m) and control strain (μ_w) under the same selection condition on days 2 and 3 of the adaptation experiment (Equation 3.2). $s_{m,c}$ was calculated as the relative change between the growth rate of the CRISPRi strain on days 2 and 3 with respect to day 1 (Equation 3.3). Positive α indicates that the CRISPRi strain adapted well, a negative α indicated that the strain adapted poorly, and α close to zero indicates that adaptation was not impacted (Fig. 3.4). The magnitude of the adaptation factor (|α|) is thus used as a gauge of the absolute impact of a gene perturbation on adaptation, as the perturbation could impact adaptation either positively or negatively.

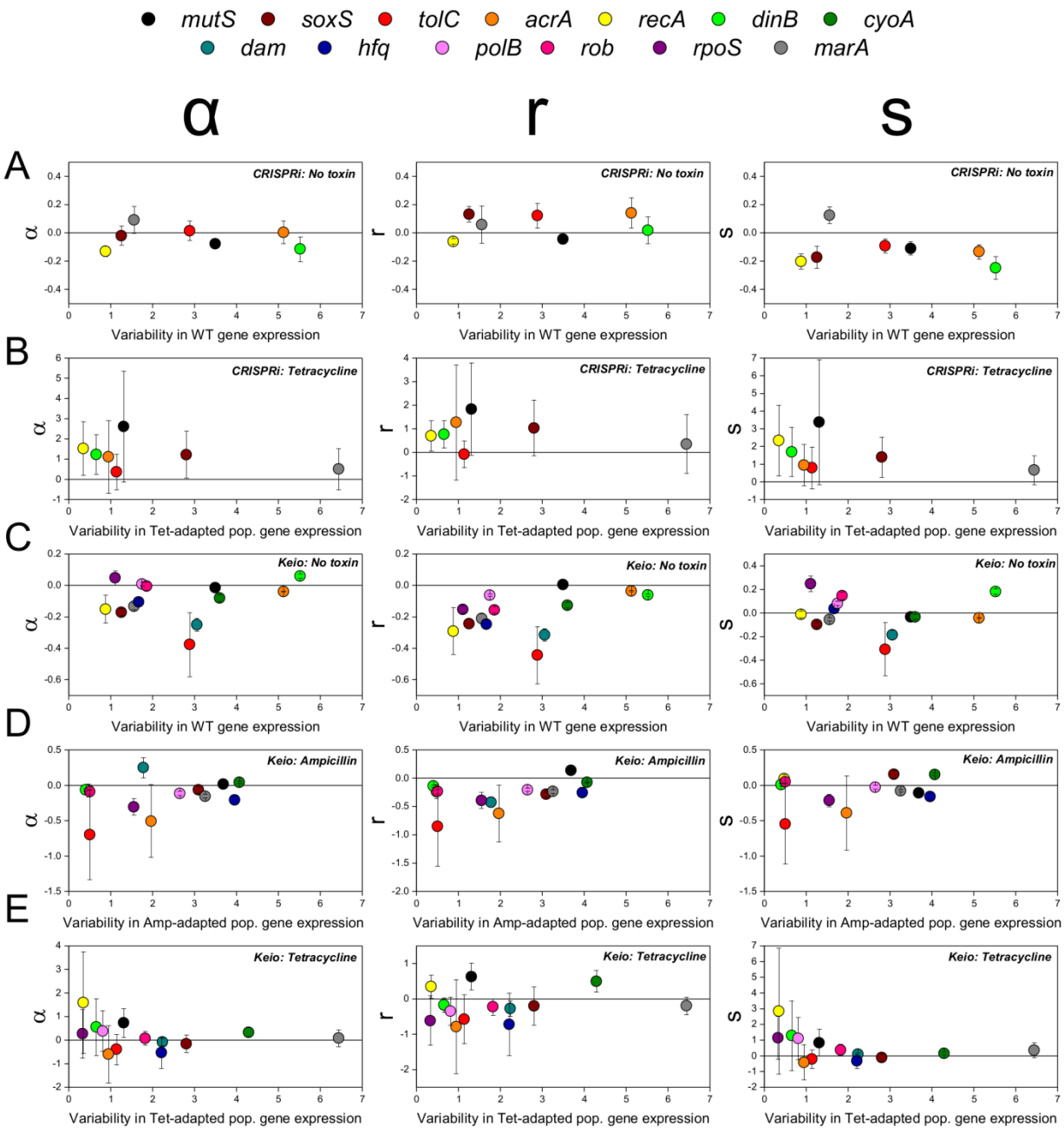


Figure 3.4 Adaptation factor vs. gene expression variability

Comparison of mRNA expression variability (x-axis) to adaptation factor (α, in first column) or the components of adaptation factor (r in second column and s in third column) for corresponding CRISPRi repression strains (A-B) and Keio single gene knockout strains (C-E). Equations for α, r, and s are presented in the text. Each point represents one gene/strain pair. Variability is the inter-population range in gene expression ($-\Delta C_{q,avg}$) between either wild-type (A, C), tetracycline-adapted (B, E), or ampicillin-adapted populations (D).

$$\alpha_{m,c} = 0.5(r_{m,c} + s_{m,c}) \tag{3.1}$$

$$r_{m,c} = 0.5\left[\left(\frac{\mu_{m,c,day2} - \mu_{w,c,day2}}{\mu_{w,c,day2}}\right) + \left(\frac{\mu_{m,c,day3} - \mu_{w,c,day3}}{\mu_{w,c,day3}}\right)\right] \tag{3.2}$$

$$s_{m,c} = 0.5\left[\left(\frac{\mu_{m,c,day2} - \mu_{m,c,day1}}{\mu_{m,c,day1}}\right) + \left(\frac{\mu_{m,c,day3} - \mu_{m,c,day1}}{\mu_{m,c,day1}}\right)\right] \tag{3.3}$$

We compared the |α| for the CRISPRi strains to the inter-population range of gene expression from adapted populations by calculating linear fits and Pearson correlation coefficients. Using an F-test, we show that the fits for unadapted and adapted conditions are statistically different (Fig. 3.5A).

A

Strain	Condition	Slope	Pearson's r	Prob > F
CRISPRi	Unadapted	-0.005±0.012	-0.164	$P = 7x10^{-4}$
	Adapted	-0.112±0.081	-0.526	
KEIO	Unadapted	-0.008±0.021	-0.110	$P = 0.080$
	Adapted	-0.092±0.040	-0.420	

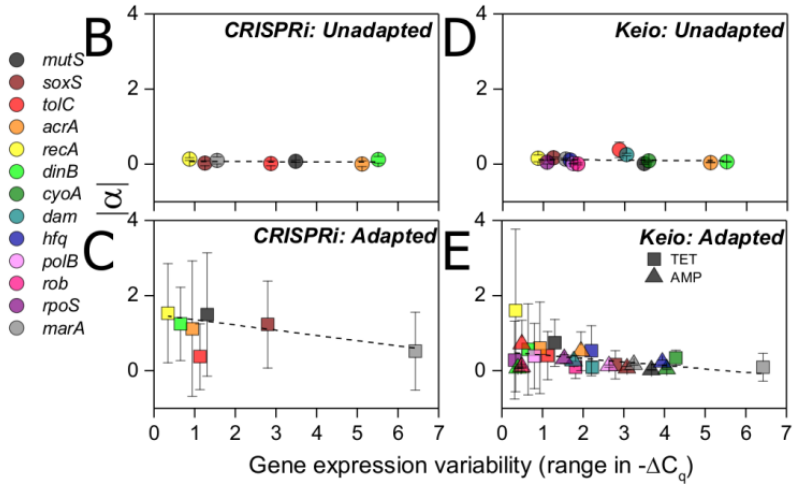


Figure 3.5 Relationship between adaptive potential and gene expression variability in modified strains of *E. coli*

(A) Parameters from linear fits shown as dashed lines in B-E, including slope with SE. (B-E) Comparison of gene expression variability (inter-population range in –ΔCq,avg) to the magnitude of the adaptation factor (|α|) of corresponding CRISPRi (B, C) or single gene knockout (D, E) strains. Y-error bars are the SD of n=3 (CRISPRi) or n=4 (knockout) biological replicates either grown without (B, D) or with (C, E) antibiotic selection pressure.

Square and triangle symbols in denote selection pressure of tetracycline and ampicillin, respectively. Each point represents one gene/strain pair.

When the strains are not under stress (unadapted condition), alpha values are all close to zero, indicating that mutant strains do not have inherently compromised fitness (Fig. 3.5B). We also find no definite correlation between |α| and gene expression variability in the unadapted condition (Fig. 3.5B, Pearson's r = -0.16 with P = 0.73). Contrastingly, CRISPRi strains adapted to tetracycline demonstrated a negative correlation (Fig. 3.5C, Pearson's r = -0.53 with P = 0.23), which hints that genes with low expression variability during adaptation to tetracycline have greater impact on adaptation to tetracycline pressure. To further scrutinize the relationship between adaptation and gene expression variability, we extended the analysis to establish whether similar correlations were present in related *E. coli* strains.

Mutant single gene knockout strains derived from *E. coli* BW25113[138] strains were grown in minimal medium with (4 μg/mL ampicillin or 2 μg/mL tetracycline) or without antibiotics to evaluate the improvement in growth rate for four replicate populations over three days of propagation. We calculated α for the mutant strains using Equations 3.1-3, with *E. coli* BW25113 as the control strain. Similar to the CRISPRi strains, the linear fits for unadapted and adapted strains were significantly different (Fig. 3.5A). Again, when strains are grown without stress, alpha values are close to zero and there is no clear relationship between adaptation and gene expression variability (Fig. 3.5D, Pearson's r = 0.11 with P = 0.72). Notably, the magnitude of the adaption factor and amount of gene expression variability are significantly negatively correlated for adapted knockout strains (Fig. 3.5E, Pearson's r = -0.42 with P = 0.03), once more indicating that genes with a

smaller range of expression have a greater impact on adaptive capability. The significance of negative correlation holds even if either of the points at the extreme end of the distribution is removed: Δ*marA* in tetracycline (r = -0.42 with *P* = 0.04) or Δ*recA* in tetracycline (r = -0.40 with *P* = 0.05). Moreover, we observe that the replicate populations with genes lower in expression variability knocked out tended to demonstrate greater heterogeneity in growth resulting in greater variability (SD) in |α| (Fig. 3.5E). Indeed, gene expression variability was also found to be significantly negatively correlated with the inter-population variability (SD) in |α| for adapted knockout strains (Pearson's r = -0.46, *P* = 0.02), but not in unadapted strains (Pearson's r = -0.16, *P* = 0.60).

3.3.4 Transcriptome-level signatures based on differential gene expression variability.

Given that gene expression variability in *E. coli* stress response genes correlates with the ability of the bacteria to adapt, we hypothesized that transcriptome-level shifts in gene expression would be present when comparing unadapted and adapted populations (Fig. 3.6A). These shifts would point to genes with significant differential variability (DV), which could potentially serve as relevant biological levers to control bacterial adaptation to stress conditions. To quantify this, we took two of the populations of *E. coli* adapted to ampicillin, tetracycline, and n-butanol for 11-14 days, and harvested total cell RNA. RNA was also harvested from wildtype *E. coli* and n-hexane exposed but unadapted populations. The RNA was subsequently sequenced in libraries on an Illumina HiSeq 2000.

To quantify gene expression variability, we compared the variability in normalized transcript abundance (coefficient of variation, CV=σ/μ in FPKM) across the four unadapted (two hexane and two wildtype populations) to that across all six adapted populations. Overall, there was a significant shift (*P*=8.6e-11) towards increased gene expression variability across adapted versus unadapted populations (Fig. 3.6B). Hierarchical clustering further underscores the shifts in expression variability at the transcriptome level and underscores sets of genes with similar trends in inter-population expression variability (Fig. 3.6C), including genes with higher variability (red) as well as lower variability (blue). To better comprehend the impacts of the variability shifts observable in this transcriptome-level heatmap, we calculated the ΔCV for each gene between unadapted and adapted populations and plotted in Fig. 3.6D.

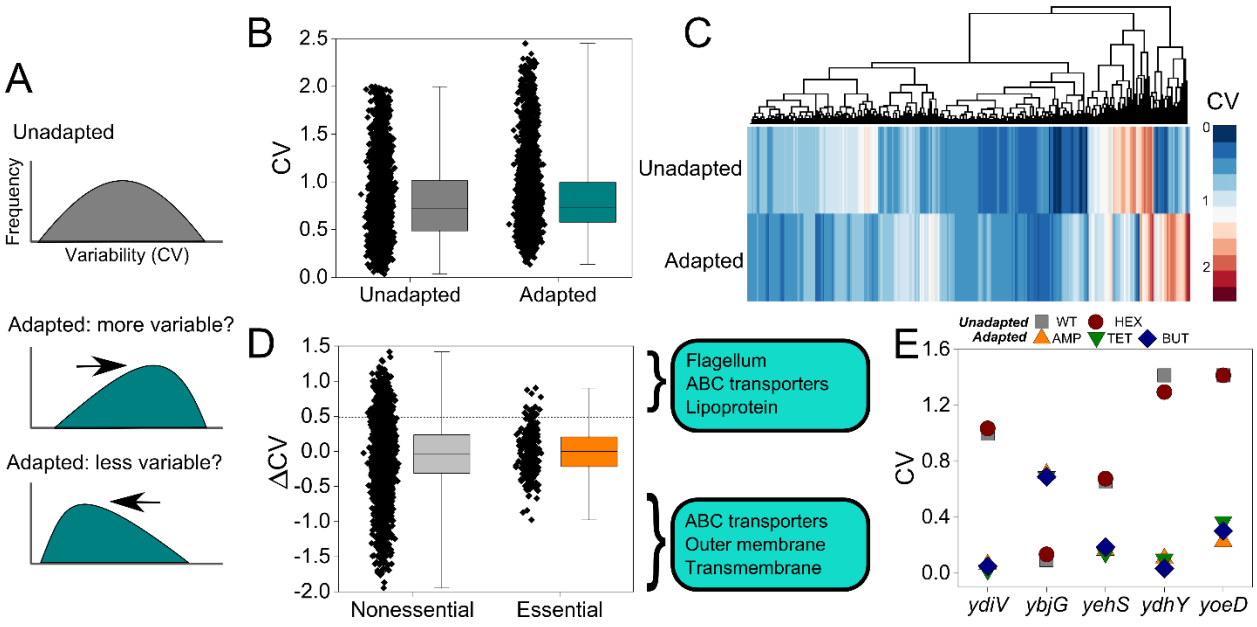


Figure 3.6 Shifts in gene expression variability are present during bacterial adaptation

(A) Hypothetical distribution in inter-population gene expression variability. If unadapted samples possess a certain distribution, we predict that shifts in variability will occur in

adapted populations. (B) Distribution of variability (CV in FPKM) in gene expression across 4181 genes in unadapted and adapted samples. For box plots in (B) and (D), all data points are shown for each condition. Box plots display the interquartile range and median for the corresponding data. Whiskers on box plots show the minimum and maximum values. (C) Hierarchal clustering by gene expression variability reveals clusters of genes (on vertical axis) with higher and lower variability in unadapted vs. adapted bacterial populations. (D) Shifts in gene expression variability in nonessential and essential genes. Shifts are defined as $\Delta CV = CV_{unadapted} - CV_{adapted}$. For $\Delta CV < 0$, the gene has higher expression variability in adapted populations. For $\Delta CV > 0$, the gene has lower variability in adapted populations. The three most enriched gene ontologies are displayed for the 10% of genes with highest and lowest ΔCV (10th and 90th percentiles in ΔCV for all genes are marked with horizontal dashed lines). (E) The CV across duplicates for five genes with significantly different expression variability in adapted vs. unadapted populations.

When examining variability per the essentiality reported in the PEC database (http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp), non-essential genes demonstrated larger shifts in variability than essential genes (Fig. 3.6D) (*P*=9.2e-11), suggesting that essential genes are more tightly regulated than non-essential genes. We located enriched gene ontologies in genes (essential and nonessential) with the largest (10^{th} and 90^{th} percentiles) shifts in variability for unadapted versus adapted strains (Fig. 3.6D, Appendix A) via DAVID v. 6.8[139]. The three most enriched classes for genes exhibiting lower variability upon adaptation ($\Delta CV>0$) were flagellum (enriched 17-fold, *P*=1.5e-12), ABC transporters (47-fold, *P*=4.3e-18), and lipoprotein (37-fold, *P*=8.1e-14). The three most enriched classes for genes with higher variability upon adaptation ($\Delta CV<0$) were associated with ABC transport (enriched 49-fold, *P*=1.5e-16), outer membrane (41-fold, *P*=3.2e-9), and transmembrane (4.1-fold, *P*=3.3e-52). We looked deeper into the differences between the set of ten ABC transporters that had higher variability upon adaptation and the set of eleven ABC transporters with lower variability. Both sets encompass transport of diverse compounds, including phosphonates, antibiotics,

sulfonates, xylose, galactose, and taurine in the higher variability set and molybdenum, nickel, glutamine, zinc, glycine, proline, glycerol-3-phosphate, peptides, and sulfates in the lower variability set. Iron transport appeared in both sets; the iron import gene *afuC* was in the high variability group, while the iron export gene *fetA* was in the lower variability set. Four of the genes in the lower variability set (*nikD*, *glnQ*, *nikE*, and *ddpF*) are regulated by the nitrogen-associated sigma factor σ^N. Transcription factors regulating genes in the low variability set only include NarL, NikR, Zur, H-NS, and NtrC. Genes regulated by σ^S, σ^{70}, CRP, FNR, Fis, IhfAB, CysB, or PhoB appear in both sets. Regulators controlling only ABC transport genes in the high variability set include σ^{32}, LsrR, Cbl, FlhCD, GalR, and GalS.

We identified a subset of five genes with significant DV in gene expression by using t-tests and controlling the false discovery rate with Benjamini and Hochberg's adjustment[140] (Fig. 3.6E). Notably, though the overall trend is towards increased variability, only *ybjG* showed significantly higher variability in adapted populations, whereas the remaining four DV genes (*ydiV*, *yehS*, *ydhY*, and *yoeD*) displayed significantly lower variability across adapted populations. Prior research supports that some of the DV genes influence resistance or stress response; for instance, *ybjG* is a putative bacitracin resistance gene[141], the motility-associated gene *ydiV*[142] is regulated by the membrane stress sigma factor σ^E, and the general stress sigma factor σ^S regulates the predicted oxidoreductase *ydhY*[143,144].

The functions of the conserved protein *yehS* and the pseudogene *yoeD* are unknown. We applied NCBI's BLAST[145] to examine potential roles for these genes. *yoeD* has more than 95% alignment with transposases in *E. coli* and *Shigella* species,

Citrobacter rodentium, and *Proteus vulgaris. yehS* has 100% alignment with hypothetical proteins in various *E. coli* and *Shigella* species. Interestingly, a 436 bp portion of the gene has 78% local alignment with DNA polymerase I in *Klebsiella oxytoca* KONIH1.

3.3.5 Application of CRISPRi to assess impact of target genes on adaptation.

The DE and DV genes presented here may offer interesting targets for attempts to manipulate adaptive mechanisms. To validate our approach for identifying key players and demonstrate that these genes are involved with adaptation, we applied CRISPRi (Fig. 3.7A). We targeted four of DV genes (*yoeD*, *ybjG*, *yehS*, *ydiV*), as well as five genes that were determined to be differentially expressed (DE) in one particular direction (*fiu, fliA*, *tar*, *wzc*, *yjjZ*). As a control, we included a plasmid expressing an sgRNA targeting a red-fluorescent protein (RFP) not present in *E. coli* MG1655. Colonies from each strain (hereby referred to by to the gene that is targeted and a "-i" to convey interference) were subjected to a range of antibiotic concentrations, and a visual Resazurin assay[146] was used to ascertain the MIC for each colony (Fig. 3.7B-D). The MIC, as well as the degree of heterogeneity introduced in the MIC, were used as indicators for each gene's involvement in adaptive resistance.

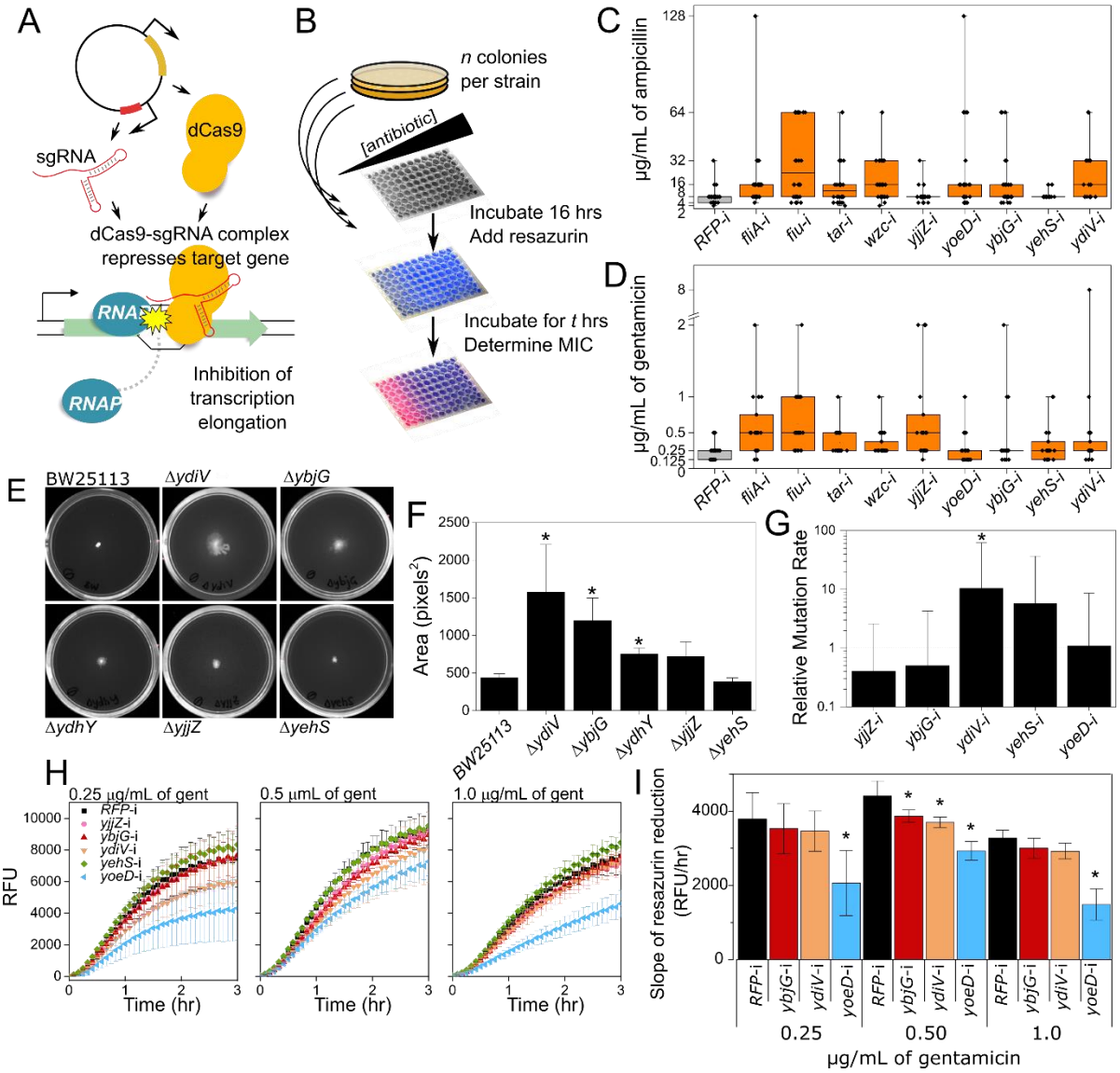


Figure 3.7 Synthetic perturbation of differentially expressed and differentially variable genes

(A) CRISPRi is used to repress gene expression by blocking progression of RNA-polymerase (RNAP) at a site specified by the sequence of the sgRNA. (B) Minimum inhibitory concentration (MIC) was determined for individual colonies from each CRISPRi strain. Colonies were grown for 16 hours in a range of antibiotic concentrations, and MIC was determined visually through a Resazurin assay. MIC for ampicillin (C) and gentamicin (D) is shown for individual colonies from each CRISPRi strain. Box plots show the interquartile range. The median is marked with a horizontal line. Whiskers demarcate min/max value. Individual data points are overlaid on the box plots. n=19-50 colonies per strain. (E) Representative plates from swarming motility assay, for E. coli BW25113 wild-type and five knockout strains after 48 hours of growth. (F) Average area from swarming motility assay. Error bars represent the SD across n=3 replicates per strain. (G) Relative

mutation rates for CRISPRi strains (rate of strain/rate of RFP-i control). Error bars represent the SD, n=30 parallel cultures for each. (H) Resazurin reduction curves. RFU = Relative fluorescence units. Error bars are the SD, n=4 replicates. (I) Slopes of Resazurin reduction curves in (H). For (E), (G), and (I), asterisks indicate a result significantly different than the control ($P < 0.05$).

When establishing the MIC to ampicillin (Fig. 3.7C) or gentamicin (Fig. 3.7D), perturbation of the target genes generally resulted in increased intra-strain heterogeneity in MIC. The range in MIC increased (relative to the control) in seven out of the nine strains. As the ability to generate diversity is a hallmark of adaptation[126,147,148], this finding is a strong indicator that both DE and DV genes influence adaptive resistance. We used one-way ANOVA and Bonferroni tests to determine whether significant differences existed in the average MIC. In ampicillin (Fig. 3.7C), both *yoeD*-i (P=0.006) and *fiu*-i (P=8.6e-7) had significantly higher average MIC than the control. The MIC of *fiu*-i was also significantly higher than that of *tar*-i, *yjjZ*-i, *ybjG*-i, and *yehS*-i (P<0.03 for all), while *yoeD*-i had a higher MIC than *yjjZ*-i (P=0.02).

While none of the strains exposed to gentamicin had significantly higher average MIC, the trend of increased heterogeneity was maintained (Fig. 3.7D). Furthermore, increased heterogeneity in MIC is associated with higher frequency of clinically relevant resistances. For instance, the Clinical and Laboratory Standards Institute (CLSI) sets an ampicillin resistance breakpoint at ≥32 μg/mL of ampicillin[149]. By this standard, one of the fifty (2%) *RFP*-i colonies was resistant to ampicillin. In contrast, resistance was achieved in >20% of *fliA*-i, *fiu*-i, *wzc*-i, *yoeD*-i, and *ydiV*-i colonies. The *fliA*-i and *yoeD*-i strains demonstrated the highest range in ampicillin MIC, with colonies in each having MICs as low as 4 or as high as 128 μg/mL of ampicillin (Fig. 3.7C). The CLSI breakpoint for gentamicin resistance is ≥16 μg/mL. Although no colonies were gentamicin-resistant (Fig.

3.7D), the general trend of increased frequency of higher MIC was maintained. Fifteen percent of control *RFP*-i colonies had a MIC ≥ 0.5 μg/mL of gentamicin, while all CRISPRi strains had a higher proportion of colonies with a MIC at or above 0.5 μg/mL, including in 60% of *fliA*-i, 70% of *fiu*-i, and 55% of *yjjZ*-i colonies.

As there were significant differences between CRISPRi strains, with *yjjZ*-i and *yehS*-i especially having ampicillin MIC distributions similar to the control, we can deduce that these results are attributable to the specific perturbation of the target gene, and not merely to any disruption of normal cellular function. This data demonstrates that subtle expression changes in certain genes can impact the likelihood of survival in the presence of high levels of stress, providing a more favorable environment in which to develop heritable resistances.

3.3.6 Certain target genes impact swarming motility

As our gene ontology enrichment analysis identified many motility-associated changes in DE or DV genes, we sought to determine whether DE/DV genes of unknown function also influence adaptation through a motility-associated mechanism. We obtained gene knockout strains for the DV genes *ydiV*, *ybjG*, *ydhY*, and *yehS* and the DE gene *yjjZ*, then compared the motility to the wild-type strain *E. coli* BW25113. Fig. 3.7E shows a representative image of each strain after 48 hours of growth on semi-solid agar plates (M9 minimal media with 0.3% agar). We find significant increases in motility in the Δ*ydiV*, Δ*ybjG*, and Δ*ydhY* strains (Fig. 3.7F). Overexpression of the anti-FlhDC factor *ydiV* has been previously shown to decrease motility[142], in agreement with our findings. However, neither *ybjG* nor *ydhY* have been previously shown to influence motility. While it is not

straightforward to rationalize how changes in variability of these genes might be reflected in a phenotype, our results suggest that shifts in variability of *ydiV*, *ybjG,* and *ydhY* could lead to phenotypic heterogeneity in motility, in turn promoting survival in the presence of stress.

3.3.7 Mutation and metabolic rates in CRISPR-interference strains

Increased mutation rates could be a mechanism for higher and more heterogeneous MIC. We performed standard fluctuation tests to assess whether CRISPRi influences mutation rates. For four out of the five CRISPRi strains evaluated, we found that the mutation rates were not significantly different between CRISPRi strains and the control (Fig. 3.7G). Therefore, we can conclude that the CRISPRi system does not inherently increase mutation rates independent of the gene being targeted, and that unintentional increases in mutation rate are not the likely explanation for the phenotypic heterogeneity present in CRISPRi strains like *yoeD*-i, *ybjG*-i, and *yjjZ*-i. Interestingly, we did observe that the *ydiV*-i strain has a mutation rate significantly higher than the *RFP*-i control (10-fold higher). As mentioned, *ydiV* does have a known function as an anti-FlhDC factor. Our results suggest that it may impact mutation rates as well, though further investigation is needed to elucidate the precise mechanism by which *ydiV* contributes to adaptive resistance.

Our gene ontology enrichment analysis also revealed a differential expression in a multitude of genes associated with metabolism. Thus, the DV and DE genes of unknown function could potentially impact adaptation by promoting changes in metabolism. Resazurin dye is reduced to the florescent Resorufin through an irreversible reaction

catalyzed by dehydrogenases and dependent upon NADH availability [150]. Therefore, by adding Resazurin to cultures and monitoring the changes in fluorescence over time, high-level insight into metabolic rates within populations can be garnered. We used a Resazurin reduction assay to measure metabolic rates in CRISPRi strains subjected to a range of gentamicin concentrations (Fig. 3.7H). We find that *yoeD*-i has consistently decreased metabolic rate as measured by the slope of the Resazurin reduction curve for a range of gentamicin concentrations, including in 0.25 µg/mL (P = 0.02), 0.5 µg/mL (P = 0.0008), and 1 µg/mL (P = 0.0003) of gentamicin (Fig. 3.7I). The average metabolic rate of *yoeD*-i strains was reduced by 46% (compare to the control) in 0.25 µg/mL of gentamicin, by 34% in 0.5 µg/mL of gentamicin, and by 55% in 1 µg/mL of gentamicin. None of the other strains had metabolic rates reduced to such an extent or reduced in more than one concentration, though *ybjG*-i and *ydiV*-i had slightly reduced rates in 1 µg/mL of gentamicin (12% reduction, P = 0.05 and 16% reduction, P = 0.02 respectively).

3.4 Discussion

In the face of pervasive heterogeneity, it is challenging to discern key genes that enable the bacteria to rapidly respond to environmental exposure. Here, we sought to ascertain unique gene expression patterns that emerge during bacterial exposure to antibiotics and biofuels. We did this through both an *a priori* approach in which we analyzed the expression profiles of 14 genes highly correlated with *E. coli* stress response, and a *de novo* approach in which we performed whole transcriptome RNA sequencing of *E. coli,* applying these approaches both before and after stress exposure.

We found that stress response genes demonstrated varying degrees of inter-population gene expression variability that depended highly on the nature of stress exposure. Furthermore, the multiple intersections of regulation in stress response networks render their corresponding gene expression patterns difficult to interpret. However, while high expression variability was observed for some genes, several genes were expressed at a consistent level across adapted populations. Furthermore, clustering based on inter-population gene expression variability enabled us to separate adapted populations from unadapted populations, indicating that the degree of variability has biological relevance. The implication of inter-population gene expression variability on adaptive resistance was further analyzed by adapting CRISPRi strains and single gene knockout mutants to antibiotics. Significantly different trends were found to relate adaptation and gene expression variability in adapted versus unadapted populations. Especially in single gene knock-out mutants, we noted a negative correlation, whereby knocking out stress response genes with low inter-population variability in expression (e.g. *tolC*, *recA*, *dinB*) was more likely to impact adaptation (either positively or negatively), while knocking out genes with high inter-population variability in expression (e.g. *marA*, *soxS*, and *cyoA*) had less impact on adaptation.

We postulate that a range of mRNA levels is selected for during adaptation, and as such, the observed expression variability provides a measure of a gene's participation in bacterial adaptive processes. Genes with low and high variability in expression likely play different roles in long-term population survival. The biological relevance of differential gene expression variability has been proposed by others. Studies in complex organisms including *Homo sapiens* report similar variability, or "dynamism" at the tissue level that

contributes to various disease states[151,152]. Recently, differential gene expression variability was observed during adaptive resistance to antisense therapeutics[97]. This work substantiates the idea that bacterial adaptation is enabled not only by changes in gene expression levels, but also by shifts in gene expression variability

The transcriptome patterns in heterogeneous adapted and unadapted bacterial populations prove new evidence of key genes and pathways involved in bacterial adaptive responses to antibiotics and biofuels. While others have used mutant library selection approaches to detect genes which convey specific tolerances or resistances[135,153], only transcriptome profiling allows for the detection of subtle and simultaneous changes across multiple genes. Ascertaining general signatures of adaptation is not trivial, due to immense potential for heterogeneity in gene expression during adaptation[127,148,154]. Next-generation sequencing technologies will help to ease this difficulty[155,156].

Our whole-transcriptome variability analysis is in line with those performed in eukaryotic systems; essential genes experienced lower magnitudes of variability shifts upon adaptation when compared to nonessential genes. In this study, the observation that variability shifts occur during adaptation is also consistent with previous studies in yeast, which have demonstrated that expression “noise” is a selectable trait[157]. We find an overall shift toward increased variability in adapted versus unadapted populations. This could be attributed to the fact that different gene expression states are being selected for across divergent populations, but could also be due to intrinsic regulation of an adaptive response, considering that stress response genes have been found to tend toward higher variability in mice and yeast[158,159]. In this study, five genes with significantly different gene expression variability were located, four out of five of which had decreased

variability upon adaptation We have previously suggested that shifts toward lower variability may impart evidence of involvement with adaptation[16], and the transcriptome-level validation here implies that gene expression variability is tunable in bacteria as well as eukaryotic systems

The DE and DV genes suggest the importance of changes in motility, metabolism, membrane structure, and transport during adaptation to diverse conditions. When expression of these genes was perturbed with CRISPRi, we found increased prevalence of higher MIC as well as larger heterogeneity in MIC in both DE and DV gene targets. Indeed, many of the perturbed strains had MIC profiles more closely resembling those associated with end-point adapted populations[16]. We briefly investigated the mechanisms by which the DV genes of unknown function could be contributing to adaptive resistance. While most CRISPRi strains had mutation rates like that of the control strain, perturbation of *ydiV* increased mutation rates approximately 10-fold, a function that has not been previously attributed to this motility-associated gene. Two other DV genes, *ybjG* and *ydhY*, also appear to impact motility, providing further evidence to support that regulation of motility is important to a multitude of stress-response pathways[160]. Finally, we observed that perturbation of *yoeD* influences metabolic rates in a range of gentamicin concentrations, suggesting a metabolic or global regulatory role for this gene of unknown function.

Taken together, our results support the existence of a non-genetic basis for adaptive resistance; subtle gene expression changes are sufficient to drive increased resistance in bacterial populations. Continued inquiry using approaches related to those presented here, and expanding to investigate additional stress conditions and bacterial

species, will only advance our understanding and ability to impede the upstream, non-genetic responses that enable the eventual emergence of novel antibiotic resistance.

3.5 Methods

3.5.1 Bacterial strains, media, and culture conditions

E. coli K-12 strain MG1655 (ATCC 700926) was used as the wild type in adaptation experiments and the host strain for harboring CRISPR-inhibition (CRISPRi)[44] synthetic constructs. Keio collection mutants and parent strain (*E. coli* BW25113, the wild-type for knock-out mutant studies) were purchased from Yale's Coli Genetic Stock Center (http://cgsc.biology.yale.edu/index.php). All adaptation cultures were grown in M9 minimal media with 0.4% glucose and toxin (ampicillin, tetracycline, n-butanol, or n-hexane) as indicated. Keio collection mutants were also grown in M9 minimal media (consisting of 5X M9 minimal media salts solution from MP Biomedicals, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$ in sterile water) with 0.4% glucose. Cells harboring CRISPRi constructs were grown in LB media (Sigma-Aldrich) supplemented with ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL) to maintain plasmid selection. Construct expression was induced via supplementation with anhydrous tetracycline (aTc) (10 ng/mL).

Colonies of MG1655 wild-type strain were grown on LB agar plates. Colonies of Keio mutants were grown on LB agar plates supplemented with kanamycin (50 μg/mL). Colonies harboring CRISPRi constructs were grown on LB agar plates supplemented with ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL).

3.5.2 Adaptation of E. coli MG1655 strains

Wild-type and n-hexane samples were obtained by growing *E. coli* MG1655 to $0.5 \leq OD_{600} \leq 1.0$ in M9 minimal media with 0.4% glucose (and with 10% v/v hexane for n-

hexane samples). OD_{600} was obtained on a Nanodrop 2000 (Thermo Scientific) using 2 µL of culture. Initially, 5 mL of overnight cultures were inoculated from three individual colonies of wild-type MG1655 strain grown at 37^0C with 225 rpm shaking in M9 minimal media with 0.4% glucose (day -1). Portions of the triplicate wild-type cultures grown to exponential phase in minimal media were stored for transcriptome analysis (day 0). We defined MIC as 2x the concentration of toxin that allows culture density of OD_{600}>0.5 after 24 hours of growth. The initial MIC of *E. coli* MG1655 was determined by splitting the day 0 wild-type MG1655 cultures in minimal media with a range of toxin concentrations. The culture growing to OD_{600}>0.5 at the highest concentration of toxin after overnight growth was selected for continued evolution. This culture was repeatedly diluted 1:5 to 1:100 (depending on culture density) to OD_{600}~0.1 in fresh media every 1-3 days into new cultures containing the toxin at the current and higher concentrations. Between splits and at least every 24 hours, media was refreshed by either diluting the culture 1:5-1:100 in new media (if $OD_{600}\geq0.5$) or spinning down cells and resuspending in new media (if $OD_{600}<0.5$) to maintain toxin pressure and nutrient supply. Evolution was continued for at least 11 days until bacterial populations were able to grow at four times the initial MIC, or until the populations exhibited no improvement in tolerance for seven continuous days. Samples for qPCR were collected on the final day of adaptation when culture at the highest toxin concentration was in exponential phase, $0.5\leq OD_{600}\leq1.0$. To store samples for RNA extraction, 0.5 mL of culture was added to 1 mL of RNAprotect Bacterial Reagent (Qiagen), incubated at room temperature for 5 minutes, and centrifuged at 4000 rpm for 10 minutes. Pellets were flash frozen in ethanol and dry ice and stored at -80^0C until RNA extraction. Freezer stocks were prepared by centrifuging 0.5 mL culture for 5 min at 4000

rpm then suspending in LB+50% glycerol. Three independent biological replicates were adapted to each toxin.

3.5.3 Adapted population growth and MIC

Adapted and wild-type populations were inoculated from glycerol stocks into M9 minimal media, grown to OD_{600}=1, diluted 1:10,000 in sterile water, and 10 µL were plated onto LB agar. After 16 hours of growth at 37^0C, 36 colonies were picked and suspended in 20 µL sterile water. For growth curves, 1 µL of each colony suspension was used to inoculate 50 µL cultures in a 384-well plate. Each colony from the wild-type and ampicillin adapted populations was grown in media with no toxin, 50 µg/mL or 100 µg/mL of ampicillin. Colonies from wild-type and tetracycline adapted populations were grown in media with no toxin, 1 µg/mL or 2 µg/mL of tetracycline. For MIC assessment of individual colonies, 1 µL of each colony suspension was spotted onto LB agar plates across a range of concentration of toxin (0, 6.25, 12.5, 25, 50, 100, 200 µg/mL of ampicillin and 0, 0.25, 0.5,1, 2, 4, 8, 16, 32 µg/mL of tetracycline). The MIC was estimated as the lowest concentration of antibiotic at which no growth was observed on solid media after 24 hours at 37^0C. All growth curves in this study were obtained with a Tecan GENios plate reader (Tecan Group Ltd.) with Magellan software version 7.2. Absorbance was read at 562 nm and 37^0C every 20 minutes, with shaking between measurements. Growth curve parameters were calculated using the program GrowthRates[161].

3.5.4 CRISPRi design and plasmid assembly

Single guide RNA (sgRNA) plasmids were derived from the RFP-targeting control plasmid pgRNA (Addgene plasmid 44251). Primers were designed to replace the 44251 plasmid's RFP-targeting sgRNA using a common reverse primer flanked with an ApaI restriction site and unique forward primers flanked with a SpeI restriction site. PCR with Phusion High-Fidelity DNA Polymerase (New England Biolabs) was used to amplify these new target sgRNA-insert DNA fragments, which were subsequently gel-purified (Zymoclean Gel DNA Recovery Kit – Zymo Research), digested with ApaI and SpeI (FastDigest Enzymes – Thermo Scientific) and PCR-purified (GeneJET PCR Purification Kit – Thermo Scientific). The 44251 plasmid backbone was also digested with ApaI and SpeI and gel purified, and T4 DNA Ligase (Thermo Scientific) was used to ligate the new sgRNA target inserts into the 44251 backbone. Ligations were transformed into chemically competent *E. coli* MG1655 cells harboring dCas9 (Addgene plasmid 44249). Plasmid minipreps were performed using Zyppy Plasmid Miniprep Kit (Zymo Research). Sequencing of final sgRNA constructs were performed for validation of correct assembly (GENEWIZ). Repression was verified with qPCR (Fig. 2.7).

3.5.5 Adaptation of E. coli MG1655 harboring CRISPRi constructs

Three individual colonies were picked from selective plates (LB agar+ampicillin+chloramphenicol). Colonies were used to inoculate 100 µL LB cultures supplemented with ampicillin, chloramphenicol, and aTc and grown to stationary phase. From each of these cultures, 2 µL were used to inoculate two 50 µL LB cultures supplemented with ampicillin, chloramphenicol and aTc (and one supplemented with 1

μg/mL of tetracycline) in a 384-well microplate. After 24 hours of growth, 2 μL of these cultures were diluted into fresh media. This process was repeated for a total of three days of growth.

3.5.6 Adaptation of E. coli BW25113 and BW25113 derived single gene knock-out mutants

Four individual colonies from each strain were picked from selective plates (LB agar+kanamycin). Colonies were suspended in 50 μL of M9 minimal media with 0.4% glucose, out of which 1 μL of each colony suspension was used to inoculate 20 μL cultures (in media only or media with 4 μg/mL of ampicillin or 2 μg/mL of tetracycline) in a 384-well microplate. The concentration of antibiotic was selected based on the MIC of 1 mL *E. coli* BW25113 cultures in M9 minimal media. After 24 hours of growth, cultures were diluted (1:40 to 1:10, depending on the absorbance recorded) into fresh media with antibiotic as indicated. Cultures were diluted twice for a total of three days of growth.

3.5.7 RNA extraction and purification for qPCR

For transcript expression from adapted populations, cell pellets from the samples at the end of each adaptation period were resuspended in TE buffer supplemented with lysozyme and proteinase K. For CRISPRi verification, overnight cultures of cells harboring CRISPR-inhibition constructs grown in M9 minimal media supplemented with 0.4% glucose, ampicillin, and chloramphenicol were inoculated with 10 ng/μL aTc for three hours, pelleted, and resuspended in TE buffer supplemented with lysozyme and proteinase K. RNA was extracted from all samples using the GeneJET RNA Purification

Kit (Thermo Scientific). DNA contamination was removed with the TURBO DNA-free kit (Ambion). RNA concentrations, A_{260}/A_{280} ratios were obtained with a Nanodrop 2000 (Thermo Scientific). Extracted and purified RNA was stored at -80^{0}C in nuclease-free water.

3.5.8 RT-qPCR

The DyNAmo SYBR Green 2-Step qRT-PCR kit (Thermo Scientific) was used to synthesize cDNA and prepare RT-qPCR reactions. In 10 µL reactions, total RNA (50-100 ng) was reverse transcribed with 2 µL M-MuLV RNase H+ reverse transcriptase (RT) and 300 ng random hexamers in a Bio-rad T100 Thermocycler. A no-RT control was included with each set of cDNA reactions. RT was carried out for 10 minutes at 25^{0}C, followed by 30 min at 37^{0}C, and 5 min at 85^{0}C. cDNA was diluted to 1 ng/µL with nuclease-free water and stored at -20^{0}C. Each qPCR reaction contained 1-2 ng template cDNA, 0.5 µM forward and 0.5 µM reverse primer, 0.2 µL 50x ROX dye, 5 µL MasterMix containing SYBR green and modified *Tbr* DNA polymerase, and nuclease-free water to 10 µL total volume. Two technical replicates were included for each sample cDNA/primer combination. The qPCR cycling program was 15 min at 95^{0}C for polymerase activation followed by 40 cycles of denaturing at 94^{0}C for 10 seconds, annealing at 55^{0}C for 30 seconds, and extension at 72^{0}C for 30 seconds. No-RT controls were included to verify that DNA contamination was negligible. Melting curves were generated after cycling was completed by holding 15 seconds at 95^{0}C, 15 seconds at 55^{0}C, then ramping to 95^{0}C and holding for 15 seconds. Data shown in this report were obtained on an Eco Illumina RT-qPCR.

Gene-specific RT-qPCR primers were designed using Primer-BLAST[162]. Sequences of open reading frames were obtained from the *E. coli* K12 MG1655 NCBI reference genome NC_000913.3. Primers were selected to be 20-22 nucleotides in length with a less than 200 nt amplicon for each primer pair. Integrated DNA Technology and Invitrogen supplied the primers, with standard desalting. Primer specificity was verified by running no-template controls and melting curves, as well as gel electrophoresis of RT-qPCR products. Primer amplification efficiency was found to be in the range of 96±10% by running standard curves with template cDNA dilutions prepared from total RNA.

3.5.9 Gene expression and gene expression variability analysis

C_q values were obtained from Eco™ Software v4.1.2.0. The average of two technical replicates (per biological replicate) was used to calculate all ΔC_q values. Cycle numbers differed by 0.3 for typical technical replicates. Due to the divergent nature of the adaptation experiments, we calculated fold change in gene expression with respect to four reference genes: *gyrA*, *hcaT*, *rrsA,* and *cysG* (Equation 3.1). Reference genes were selected from literature[163–165]. The average C_q for the four reference genes together exhibited a lower standard deviation than the average C_q across all genes, indicating that the reference genes were generally more stable than an average gene.

The heatmap and dendrograms were generated using the clustergram function with a Euclidean distance metric, unweighted average linkage function, and optimal leaf ordering[166] in MATLAB's Bioinformatics Toolbox (The Mathworks, Inc., Natick, MA). Normalized gene expression of mRNA expression was calculated per the $2^{-\Delta\Delta Cq}$ method[167], per Equation 3.4. For adapted populations, gene expression for each gene of

interest was normalized with respect to four reference genes and three wild-type strains (m=3) by taking the geometric mean of the $2^{-\Delta\Delta Cq}$ calculated for each gene of interest (goi), reference gene (ref), and wild-type strain (wt) pairing[168], as shown in Equation 3.5, where m=3. For verification of inhibition using CRISPRi constructs, normalized gene expression was calculated for individual biological triplicates as in Equation 3.4-5, using *rrsA* as a reference gene. *P*-values were obtained using two-tailed, type two t-tests.

$$\Delta C_{q,goi,avg} = \frac{1}{n}\sum_{i=1}^{n}\left(C_{q,goi} - C_{q,ref(i)}\right) \tag{3.4}$$

$$NRQ_{goi} = \left(\prod_{j=1}^{m} 2^{\Delta C_{q,goi,avg,wt(j)} - \Delta C_{q,goi,avg,sample}}\right)^{1/m} \tag{3.5}$$

3.5.10 Linear fitting

Slope and Pearson correlation coefficients for gene expression variability (range in -ΔCq) versus |α| were calculated using linear fits with no weighting (OriginPro software, v 9.1.0).

3.5.11 RNA-sequencing library preparation

Total RNA for sequencing was extracted using phenol-chloroform extraction with a TRIzol Max Bacterial RNA Isolation kit (Ambion). RNA was treated with Turbo DNA-free kit (Ambion) to remove DNA. RNA concentration and A_{260}/A_{280} ratios (>1.8) were obtained with a Nanodrop 2000 (Thermo Scientific). Ribosomal RNA treatment and library preparation were carried out at the Genomics and Microarray Core Facility (Anschutz Medical Campus, University of Colorado – Denver). Ten sequencing libraries were prepared using 80-600 ng of total RNA per sample and nonstranded Nugen kits. All

samples were sequenced in one lane of an Illumina HiSeq 2000 with 1x100bp reads, generating an average of 28.6±2.2 million reads per library.

3.5.12 Sequencing data analysis

The *E. coli* K12 MG1655 reference FASTA and gene annotation files were obtained from Ensembl, in the bacteria_22 collection (gca_000005845.2.22). The TopHat/Cufflinks workflow[169] was used to identify differentially expressed transcripts and to calculate FPKM for differential variability analysis. TopHat version 2.0.6 was applied to map reads to the Ensembl reference files, then the Cufflinks software version 2.1.1 was applied to assemble transcripts (via Cufflinks), combine transcript assembly files (via Cuffmerge), and calculate differential expression (via Cuffdiff). To account for differences in rRNA treatment during sample preparation, the mask option was used to remove the ribosomal RNA transcripts prior to calculating differential expression. Bias correction and multi-read mapping correction options were also applied. As an additional quality check, we compared expression levels obtained from Cufflinks and DESeq[170] to levels obtained from qPCR.

To investigate differential expression in each of the eight individual sample populations, calculations were performed with partial replicates: two wild-type populations considered to be biological replicates were used together to separately analyze each of the additional eight populations. Libraries were normalized by fragments per kilobase of transcript per million mapped reads (FPKM), variance was estimated with a pooled dispersion method, and genes were considered significantly DE if $P<0.05$ and $q<0.30$. To estimate differential gene expression variability (DV), the coefficient of variation (CV) for

adapted and unadapted samples was calculated from the FPKM for each gene. Genes for which FPKM values were flagged by Cufflinks to have low or high data were removed prior to analysis, as were genes with a mean FPKM=0 across any two set of duplicates, and four genes that had FPKM=0 in at least one replicate in all conditions. A two-tailed type two Student's t-test was used to compare the CV across three adapted and two unadapted growth conditions. The false discovery rate was controlled with Benjamini and Hochberg's algorithm[140]. Genes were considered significantly DV if $P<0.05$ and $q<0.30$.

Furthermore, we examined the differentially expressed genes and transcript abundances for a correlation with distance from the origin, which would indicate that minor differences in optical density at time of sampling are resulting in different growth phases that impact the results of gene expression analysis. We compared the normalized expression values from all differentially expressed genes to their chromosomal position and their absolute distance from the origin. The Pearson correlation coefficient for chromosomal position versus expression value was -0.010 with a *P*-value of 0.71. The correlation coefficient for distance from origin versus expression values was -0.023 with a *P*-value of 0.41. We also examined all genes in each library using FPKM vs chromosomal location or distance from origin, and found no significant correlation between position and transcript abundance for any of the libraries. Pearson correlation coefficients ranged from -0.015 to 0.029 with *P*-values from 0.066 to 0.95. Thus, all correlation coefficients strongly indicate that differential expression observed in this work cannot be explained by chromosomal position or proximity to the origin.

Bowtie 2 version 2.0.2[171] was used to index the reference genome and generate alignment (sam) files using end-to-end alignment mode and default scoring from each

FASTQ file. Alignment sam files were converted to sorted bam files using SAMtools version 0.1.18[172]. Indels and single nucleotide polymorphisms (SNPs) were called using the Genome Analysis Toolkit version 2.4-9[173]. Variants were called from sorted bam files from which PCR duplicates had been removed with SAMtools. SNP calls were filtered by quality by depth (QD<2.0), mapping quality (MQ<40.0, MappingQualityRankSum< -12.5), strand bias (FS>60), and position of alternate allele in the read (ReadPosRankSum< -2.0). Indel calls were filtered per quality by depth (QD<2.0), strand bias (FS>60), and position of alternate allele in the read (ReadPosRankSum< -2.0). The ReadPosRankSum requirement was made more stringent after observing that many of the false positives called were located near the end of reads. A custom Python script was used to add annotations (type of mutation, gene affected, synonymous or non-synonymous, amino acid change) in comparison to the Ensembl reference and gene annotation files. A custom MATLAB script was used to remove variants that exactly matched any variant in wild-type samples, and find overlaps between populations. The Integrative Genomics Viewer[174] (IGV) was used to visualize all variants that passed the filter in DE genes, DV genes, and transcription factors regulating DE/DV genes. In many cases, false variants passed the filtering stage (e.g. variant in a minority of reads, and only located at the end of reads).

3.5.13 CRISPRi strain construction

All CRISPRi plasmids perturbing target gene expression were derived from pRFP-i. Single guide RNA (sgRNA) targeting sequences were 20 nt long, and located immediately downstream of an NGG protospacer adjacent motif (PAM) near the 5’ end of the target open reading frame. This was accomplished by amplifying new sgRNA targets

via PCR using Phusion® High-Fidelity DNA Polymerase (New England Biolabs). Primers were designed and ordered from Invitrogen™ Custom DNA Oligos (Thermo Scientific). A common reverse primer was used, with unique forward primers to replace the 20 nt RFP-complementary sequence with information dictating the new sgRNA sequence. The resulting 124 nt fragments and pgRNA-bacteria plasmid backbone were digested with ApaI and SpeI, then ligated together with T4 DNA ligase. The recovered pgRNA-bacteria plasmids (now with new sgRNA targets) and the pRFP-i plasmid were digested with XhoI and AvrII. The 498 bp new sgRNA inserts and the 6720 bp pRFP-i backbone were gel-extracted and ligated together as previously outlined. Overnight ligations were electroporated into NEB 10-β electrocompetent cells. The plasmid recovery, confirmation, and transfer to *E. coli* MG1655 were analogous to the process outlined above.

For verification of repression, three colonies from each strain were picked from LB agar+Cm plates and inoculated into 1 mL minimal media with Cm for 16 hours at 37^0C. Cultures were diluted 1:20 into new media with 10-40 ng/mL anhydrous tetracycline (aTc) and grown for 6 hours, then samples were harvested and stored in two volumes of RNAprotect Bacterial Reagent (Qiagen). RNA was extracted with a GeneJET RNA Purification Kit (Thermo Scientific). DNA contamination was removed with a TURBO DNA-free kit (Ambion). Reverse transcription was performed with a Dynamo cDNA synthesis kit (Life Technologies) and random hexamers. qPCR reactions were prepared with Maxima SYBR Green qPCR Master Mix. Each reaction contained 1 ng of cDNA template, 0.3 μM of each gene-specific primer, 12.5 μL of master mix with SYBR green dye and Maxima Hot Start *Taq* Polymerase, 10 nM ROX dye, and water to 25 μL. The cycling protocol was a 10 min initial denaturation at 95^0C, then 40 cycles of denaturation

(95^{0}C for 15 s), annealing (55^{0}C for 30 s), and extension (72^{0}C for 30 s). qPCR was run on a QuantStudio 6 Real-Time PCR machine in the CU Core Sequencing Facility. Melt curves and no template controls were run for each primer pair to verify specificity. Standard curves were run with cDNA dilutions to check primer efficiency. No RT controls were run to confirm negligible DNA contamination. Two technical replicates were run for each of three biological replicates. Normalized gene expression was calculated according to the $2^{-\Delta\Delta Cq}$ method[167], using *rrsA* and *gyrA* as reference genes, and *E. coli* carrying a CRISPRi plasmid targeting RFP expression as a control strain (Fig. 3.10).

3.5.14 Minimum inhibitory concentration (MIC) for CRISPRi strains

E. coli MG1655 colonies harboring CRISPRi plasmids were picked from LB agar+chloramphenicol plates and suspended in 200 µL of M9 minimal media. 25 µL of each colony suspension was added to 25 µL of media to result in 50 µL cultures with 40 ng/mL of aTc, 25 µg/mL of chloramphenicol, and a range of antibiotic concentrations (0.125-4 µg/mL of gentamicin and 2-64 µg/mL of ampicillin). Cultures were propagated at 37°C with shaking. After 16 hours, resazurin dye was added to each well and the plates were incubated at 37°C. MIC was called for each colony as the lowest concentration of antibiotic for which there was no visible color change after 3 hours (ampicillin assay) or 24 hours (gentamicin assay). For colonies that grew at all concentrations, MIC was called as 2x the highest concentration tested. For colonies that grew at no concentrations, MIC was called as the lowest concentration tested. The number of colonies analyzed for each strain ranged between 19 and 50, with the average at 30 colonies for ampicillin tests and 20 colonies for gentamicin tests.

3.5.15 Fluctuation tests

Mutation rates were estimated with fluctuation tests using the method of Luria and Delbruck[175]. Single colonies of the control and CRISPRi strains from LB agar plates were used to inoculate 1 mL cultures in LB, which were grown for 16 hours at 37ºC. Cultures were diluted to the same OD_{600} per measurements on a Nanodrop 2000. Normalized cultures were each diluted 1:10,000 into thirty parallel 100 µL cultures of M9 minimal media with 40 ng/mL of aTc and 25 µg/mL of chloramphenicol. Cultures were grown for 24 hours at 37ºC. To determine viable cell counts, 2 µL of each of the 30 cultures per strain were pooled and dilutions were plated on LB agar for CFU analysis. The remaining 98 µL of each culture were plated on LB agar with 100 µg/mL of rifampicin. Colonies on rifampicin plates were counted after 48 hours. Mutation rates and 95% confidence intervals were determined via the Ma-Sandri-Sarkar maximum-likelihood method[176], implemented by the FALCOR web tool[177]. Significance was assessed with Student's t-tests using the mutation rates and confidence intervals calculated by FALCOR.

3.5.16 Swarming motility assay

We performed swarming motility assays on semi-solid plates (M9 minimal media with 0.4% glucose and 0.3% agar). We opted to use Keio collection strains[138] and *E. coli* BW25113 obtained from the Coli Genetic Stock Center (http://cgsc.biology.yale.edu/index.php), since these strains would not require aTc induction for 48 continuous hours to ensure a gene perturbation (as the CRISPRi strains would). We picked 5 colonies for each strain from LB agar plates and resuspended in 20 µL of sterile water. Plates were poured and dried for 30 minutes, then 1 µL of each colony

suspension was stabbed into the center of each small plate. Three replicates were plated for each strain. Plates were incubated at 37°C for 48 hours. We photographed the plates using a Gel Doc EZ (Bio-Rad, Hercules, CA), and measured the area of the colonies using a custom pipeline in CellProfiler[178]. Images were manually cropped, then colonies were detected using the IdentifyPrimaryObjects analysis module, with a global thresholding strategy and robust background thresholding method. Colony area was measured with the MeasureObjectSizeShape analysis module.

3.5.17 Resazurin metabolic rate assay

Overnight cultures in LB were diluted 1:100 into microplate wells with 40 µL LB, 40 ng/mL of aTc, 25 µg/mL of chloramphenicol, and a range of concentrations of gentamicin (up to 2 µg/mL of gentamicin). Four biological replicates were included for each strain. The plate was incubated for 20 hours at 37°C with 225 rpm shaking, then 4 µL of 10x resazurin was added to each well. Changes in fluorescence were monitored in a Tecan Genios (excitation 485 nm, emission 610 nm) in five-minute intervals. The slope of the curve was determined using a custom MATLAB script. The most linear slope for each replicate was determined using a sliding window of 5 time points and an R^2 value (minimum R^2 value for any replicate was 0.98). Slopes were averaged across the replicates, and two-tailed t-tests were used to compare the slope of the RFP-i control to that of each CRISPRi strain.

3.5.18 Additional software tools and statistical analysis

The ‘pca’ function from MATLAB version R2014b (The Mathworks, Inc., Natick, MA) was used for principal component analysis (PCA). Correlation coefficients and PCA were performed on FPKM generated through the Cufflinks pipeline. Genes flagged as having low or high data in any one sample were removed from all samples prior to analysis.

The ‘clustergram’ function from MATLAB’s Bioinformatics toolbox was used to perform hierarchical clustering. Dendrograms were built with a Euclidean distance metric, optimal leaf ordering, and average linkage function.

Box plots were generated and other statistical analysis were performed with OriginPro 9.1 software (OriginLab Corporation, Northampton, MA). To compare variability in unadapted and adapted populations, the CV for each gene across wild-type and n-hexane populations was compared to the CV for each gene across ampicillin, tetracycline, and n-butanol adapted populations. Significance was calculated with a two-tailed, type two t-test. To compare variability shifts in essential versus non-essential genes, the $|\Delta CV|$ was calculated as $|CV_{unadapted} - CV_{adapted}|$. A two-tailed, type two t-test was used to compare the $|\Delta CV|$ for 288 essential genes to that for 3828 non-essential genes. Genes not in the PEC database (http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp) were excluded from this analysis.

The coefficient of variation for unadapted and adapted samples is the mean of FPKM across all unadapted or adapted populations respectively divided by the standard deviations across the same samples.

To examine the function of unknown genes, nucleotide BLAST was executed through NCBI's web interface at http://blast.ncbi.nlm.nih.gov/. We compared sequences from *E. coli* K12 MG1655 genes to the NCBI Chromosome database with the megablast algorithm (optimized for highly similar sequences).

3.5.19 Data accession

Data have been deposited in NCBI's Sequence Read Archive (accession SRP069322).

3.6 Author Contributions

K.E.E. and A.C. conceived of these studies and were the primary authors of the corresponding papers. P.B.O. adapted relevant text from these papers, created the CRISPRi constructs, and performed growth experiments in Fig. 3.5B-C. K.E.E. Performed all other experiments and analysis.

Chapter 4: CRISPR gene perturbations provide insights for improving bacteria biofuel tolerance

Otoupal, P. B. & Chatterjee, A. *Submitted*

4.1 Abstract

Economically viable biofuel production is often limited by low levels of microbial tolerance to high biofuel concentrations. Here we demonstrate the first application of deactivated CRISPR perturbations of gene expression to improve *E. coli* biofuel tolerance. We construct a library of 31 unique CRISPR inhibitions and activations of gene expression in *E. coli* and explore their impacts on growth during ten days of exposure to n-butanol and n-hexane. We show that perturbation of metabolism and membrane-related genes induces the greatest impacts on growth in n-butanol, as does perturbation of redox-related genes in n-hexanes. We identify uncharacterized genes *yjjZ* and *yehS* with strong potential for improving tolerance to both biofuels. Perturbations demonstrated significant temporal dependencies, suggesting that rationally designing time-sensitive gene circuits can optimize tolerance. We also introduce a sgRNA-specific hyper-mutator phenotype (~2600-fold increase) into our perturbation strains using error-prone Pol1. We show that despite this change, strains exhibited similar growth phenotypes in n-butanol as before, demonstrating the robustness of CRISPR perturbations during prolonged use. Collectively, these results demonstrate the potential of CRISPR manipulation of gene expression for improving biofuel tolerance and provides constructive starting points for optimization of biofuel producing microorganisms.

4.2 Introduction

Bacteria have long been investigated for their ability to produce renewable, biologically-derived replacements for petroleum-based fuels such as gasoline. Microbially produced biofuels have a promising future[7,8], with particular interest in straight-chain carbon alcohols[179] and alkanes[180]. Despite their potential, biofuels represent only ~2% of total transportation-based energy consumption[181], primarily due to their low economic competitiveness. This is limited to a large degree by the inherent toxicity such products exhibit to their hosts[182].

One particularly interesting biofuel is n-butanol due to its high energy density, low volatility, and ability to interface with our current gasoline-based infrastructure[183,184]. However, in a clear representation of the aforementioned tolerance issue, butanol is one of the most toxic biofuel compounds to microorganisms[185], with yields typically limited to a maximum of 2% vol/vol under optimal conditions[186,187]. Engineering improved butanol tolerance is a key limiting factor to its economic viability and remains an elusive goal[188]. Similar problems have plagued the progress of bringing other biofuels such as n-hexane to market[189].

Increasing microbial tolerance to biofuels would go a long way towards improving their economic competitiveness and remains a high-priority research goal. While many studies have explored improving the tolerance of specialized strains such as *Clostridium*[190–192] or *Synechocystis*[193,194], recent attention has turned towards importing heterologous biofuel pathways into the well-characterized and easy to use *Escherichia coli*[195–197]. In *E. coli*, n-butanol tolerance has been associated with oxidative stress response, respiration, transport, and metabolite synthesis[134,198]. While these studies have

posed promising pathways to target, the extensive knowledge established in other strains has yet to be fully translated to *E. coli*. For instance, a 20-30% increase in membrane fluidity has been associated with n-butanol exposure in *Clostridium*, suggesting that membrane related genes could also be involved in improving *E. coli n*-butanol tolerance[199,200].

A promising approach to improving tolerance is to engineer alternative gene expression states. Manipulating gene expression is an essential metabolic engineering approach that has been previously applied to increase ethanol tolerance[201], and could similarly be applied to improving tolerance towards other biofuels[202]. This has a crucial advantage over gene knockout or insertion approaches in that it can be used to fine-tune biofuel pathways so as to not waste essential resources and restrict growth[203]. Furthermore, manipulation of gene expression can be easily implemented into genetic feedback circuits for real-time pathway balancing during biofuel production[204]. However, successful manipulation of transcriptional machinery to regulate specific genes has been difficult to achieve, preventing widespread implementation of such practices[30].

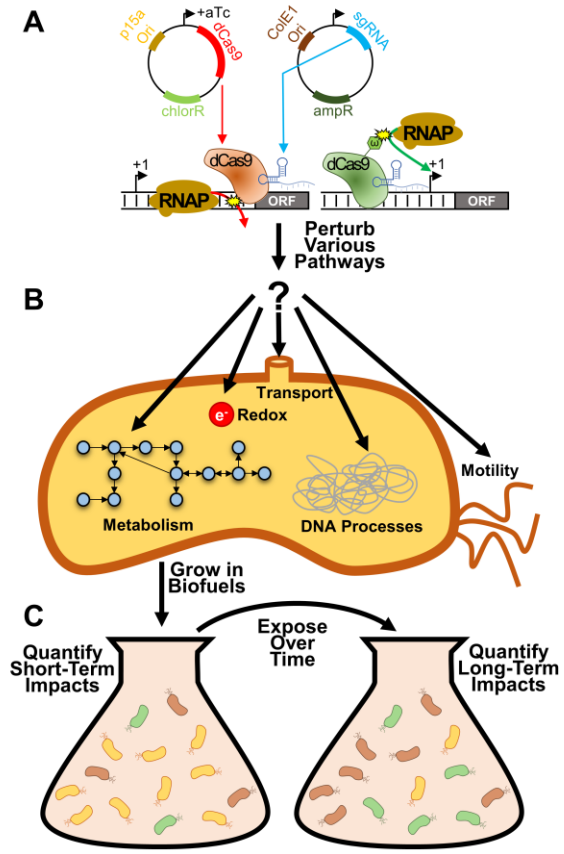
Utilizing CRISPR technology is a promising way to overcome these barriers. Deactivated versions of Cas9 have been developed to fine-tune expression patterns by inhibiting[44] or activating[45] virtually any gene in a relatively facile manner. This has sparked renewed interest in engineering gene expression to enhance biofuel production[205], as CRISPR-mediated gene modulation has the potential for fine-tuned optimization of cellular pathways[67]. Furthermore, while CRISPR-Cas9 has been applied towards the integration of heterologous genes[206,207] increasing fatty acid production[208], improving butanediol production[69], or redirecting metabolic flux[203], no work has explored the use of

deactivated CRISPR systems for improving biofuel tolerance in *E. coli*[209]. Additionally, CRISPR interference has been used to improve *Klebsiella* n-butanol production 154%, demonstrating that there is similar potential for improving *E. coli* n-butanol tolerance[68].

Here we systematically explore the growth impacts of a library of 31 CRISPR inhibitions or activations of *E. coli* gene expression during exposure to either n-butanol or n-hexane (Figure 1A). These CRISPR constructs were targeted to genes involved in a broad range of cellular processes including metabolism, redox, transport, DNA and RNA processes, and motility, as all have been implicated for their importance in determining biofuel tolerance capacity[17,19,182,185,202] (Figure 1B and 1C). We explored both inhibition and activation of gene expression, as both approaches could feasibly lead to optimization of tolerance. As growth phenotypes can be time-sensitive, we explored growth impacts over ten days of exposure to identify perturbations that impact growth phenotypes in either the short-term (one day) or long-term (ten days), as each result points to different approaches that could be implemented (Figure 1D).

Our CRISPR perturbation approach reveals a number of promising gene targets whose expression could be engineered for improved biofuel tolerance. Manipulation of metabolism-related genes, as well as membrane and periplasm related genes, appears the most promising pathways for increasing tolerance to n-butanol. Conversely, redox genes appear to be more influential in improving n-hexane tolerance. Strong temporal effects were identified under both conditions, suggesting that time-sensitive alterations of gene expression should be taken into consideration while engineering improved biofuel tolerance. We also present evidence that these perturbations are stable by artificially introducing a hyper-mutator phenotype (increasing basal mutation rates ~2600-fold)

during exposure to n-butanol[210,211]. Despite this increased mutation rate, perturbations largely demonstrate the same relative impact on growth phenotypes as before, suggesting that CRISPR perturbations maintain efficacy over prolonged periods. Together, these results demonstrate the power of CRISPR perturbations for improving biofuel tolerance.



D

Strain	Gene	Description	Pathway
dinB-i	*dinB*	DNA polymerase IV	DNA & RNA Processes
dinB-a	*dinB*	DNA polymerase IV	
fliA-i	*fliA*	RNA polymerase σ-factor 28	
mutS-i	*mutS*	Methyl-directed mismatch repair	
mutS-a	*mutS*	Methyl-directed mismatch repair	
recA-i	*recA*	DNA strand exchange & recombination	
recA-a	*recA*	DNA strand exchange & recombination	
topA-i	*topA*	DNA topoisomerase 1	
frr-i	*frr*	Ribosome recycling factor	
dfp-i	*dfp*	Coenzyme-A synthetase	Metabolism
gadA-i	*gadA*	Glutamate de-carboxylase	
wcaA-i	*wcaA*	Colanic acid biosynthesis	
wzc-i	*wzc*	Tyrosine kinase	
zwf-i	*zwf*	Glucose-6-phosphate dehydrogenase	
ampC-i	*ampC*	Serine cephalosporinase β-lactamase	Redox
marA-i	*marA*	Multiple antibiotic resistance regulator	
marA-a	*marA*	Multiple antibiotic resistance regulator	
sodB-i	*sodB*	Superoxide dismutase	
soxS-i	*soxS*	Superoxide response regulator	
soxS-a	*soxS*	Superoxide response regulator	
ybjG-i	*ybjG*	Putative bacitracin resistance	
ydhY-i	*ydhY*	Putative oxidoreductase	
acrA-i	*acrA*	AcrAB-TolC multidrug efflux pump	Transport & Motility
acrA-a	*acrA*	AcrAB-TolC multidrug efflux pump	
tolC-i	*tolC*	AcrAB-TolC multidrug efflux pump	
tolC-a	*tolC*	AcrAB-TolC multidrug efflux pump	
fiu-i	*fiu*	Outer membrane iron receptor	
ompF-a	*ompF*	Outer membrane porin 1a	
tar-i	*tar*	Methyl-accepting chemotaxis protein	
yehS-i	*yehS*	Unknown	Unknown
yjjZ-i	*yjjZ*	Unknown	
Control	*RFP*	Nonsense target of RFP	N/A

Figure 4.1 Improving bacterial tolerance to biofuels using CRISPR gene expression perturbation

(**A**) CRISPR perturbations of gene expression (both inhibition and activation) were designed for 31 *E. coli* genes and expressed using a two-plasmid system. (**B**) Strains used in this study. Whether CRISPR constructs were used to inhibit or activate gene expression are notated by -i or -a in the strain name respectively. (**C**) These perturbations were designed to disrupt expression of a variety of genes related to bacterial metabolism, redox, transport, various DNA and RNA processes, and motility. (**D**) Strains harboring these individual perturbations were exposed to biofuels (n-butanol and n-hexane) over multiple days, with the characterization of both short-term and long-term growth impacts.

4.3 Results

4.3.1 Construction of CRISPR Perturbations and Quantification of Impact on E. coli growth During No Biofuel Exposure

We first designed a diverse library of 31 CRISPR perturbations to modulate gene expression in *E. coli*. These gene targets were selected based on previous genes known to be involved in general bacterial stress response (*acrA*, *dinB*, *marA*, *mutS*, *recA*, *soxS*, and *tolC*)[19], or to be involved in central biological processes (*dfp*, *frr*, *gadA*, *topA,* and *zwf*)[182,185,202]. We also explored genes that we had previously identified to exhibit altered transcriptomic signatures during exposure to n-butanol or n-hexane[17] (*tar*, *fliA*, *fiu*, *wcaA*, *wzc*, *ybjG*, *ydhY*, *yehS*, and *ybjG*). *OmpF* has been associated with improved solvent tolerance[212], and *ampC* expression has been directly correlated to diminished *ompF* activity[213]. Finally, *sodB* overexpression has demonstrated significantly increased n-butanol tolerance[198]. These genes were selected over other genes associated with biofuel tolerance as they are not co-transcribed with other genes, or are contained in relatively small operons, so as to minimize the CRISPR perturbation's direct impact on other genes.

CRISPR inhibition constructs were designed to repress gene expression by binding within the first ~50 nt of the open reading frame, or around the +1 site of the respective promoter[214]. CRISPR activation constructs were designed to bind ~80-100 nt upstream of the +1 site of the promoter. These gene perturbations resulted in decreasing or increasing mRNA production to ~10-fold basal levels as shown in previous studies[16,17,19].

We first tested how these strains behaved during growth in plain LB, to demonstrate how perturbations impacted growth in the absence of stress (Figure 2). We

quantified the maximum growth each strain reached at the end of one day of growth (Figure 2A) and plotted the growth curves of the top and bottom three strains as ranked by maximum growth reached (Figure 2B). We also determined the growth rates and lag times of each strain in relation to the control (Figure 2C).

Out of 31 perturbation strains, only six showed significant differences in growth. Only strain DinB-a grew significantly better than the control. Five strains (RecA-i, TopA-i, WcaA-i, AmpC-i, and OmpF-a) grew significantly worse than the control, with strain OmpF-a growing noticeably poorly. Only TopA-i grew slower than the control. Two strains (MarA-a and SodB-i) exhibited a longer lag time, while five strains (RecA-a, TolC-a, AcrA-a, DinB-a, and Wzc-i) exhibited shorter lag times. All five of these strains exhibiting shorter lag times were also in the top ten strains in terms of ranked growth (Figure 2D), suggesting a slight inherent benefit to gene activation on *E. coli* growth.

Overall, these data demonstrate that growth of perturbed strains in plain LB medium was at most moderately disrupted by gene perturbations. This is most noticeable by observing the overall growth curves of the top and bottom-growing strains, which reveal similar growth trajectories relative to the control (Figure 2B). Any noted difference from the control of each perturbed strain was taken into consideration in future analysis.

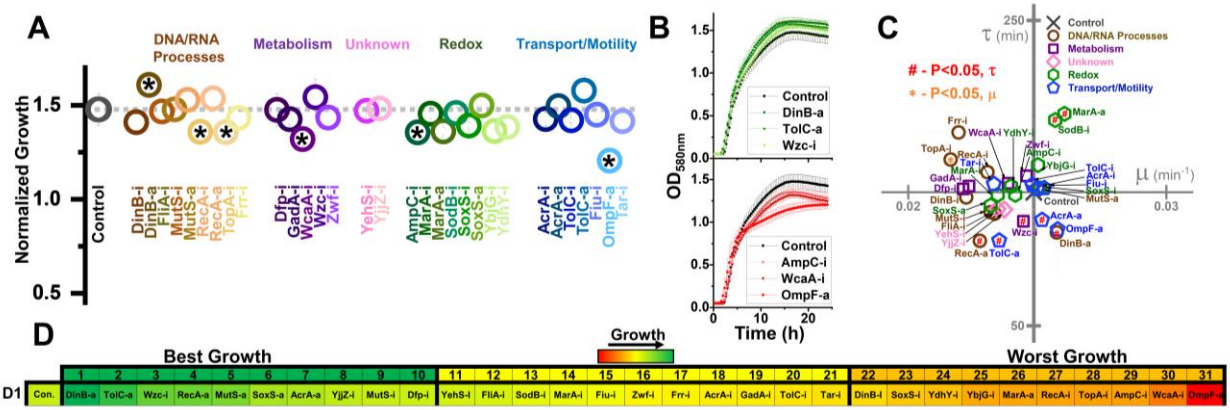


Figure 4.2 Growth of *E. coli* harboring CRISPR gene perturbations in the absence of biofuels

(**A**) Normalized growth (maximum OD/starting OD) of all strains. Strains are organized based on pathways affected by perturbation, and a dashed line extends from the control for comparison. Asterisks indicate significant differences in relation to the control ($P<0.05$). A two-tailed type II t-test was used to calculate significance relative to the control. (**B**) Growth curves of the three strains growing to the highest levels (green, top) or lowest levels (red, bottom). (**C**) Growth rates (μ) and lag times (τ) of each strain in relation to the control strain, located at the intersection of the x- and y-axes. Pathways of the affected perturbation are again indicated using symbol and color. Red # indicates significant differences in lag times, while orange * indicates significant differences in growth rates, relative to the control. (**D**) Organized rankings of strains by highest growth reached, with the color scale to indicate relative growth. The top ten and bottom ten are indicated as “best growth” and “worst growth” respectively. All error bars represent the standard deviation of four biological replicates.

4.3.2 Impact of CRISPR Perturbations on E. coli Growth During n-butanol Exposure

We next exposed our CRISPR perturbations to 0.5% vol/vol n-butanol and analyzed how each perturbation impacted growth. This experiment was performed over the course of ten days, with 1:100 dilution into fresh media at the start of each day. Growth was quantified in a microplate reader on days one, five, and ten of the experiment. We again analyzed the normalized growth of each strain on each day of the experiment (Figure 3A) and plotted the growth curves of the top three and bottom three strains on each day of the experiment (Figure 3B). We also ranked each perturbation by maximum

growth reached during each day, parsing out the top ten strains exhibiting the highest and lowest growth respectively (Figure 3C). Perturbations were compared against the nonsense targeting RFP control strain.

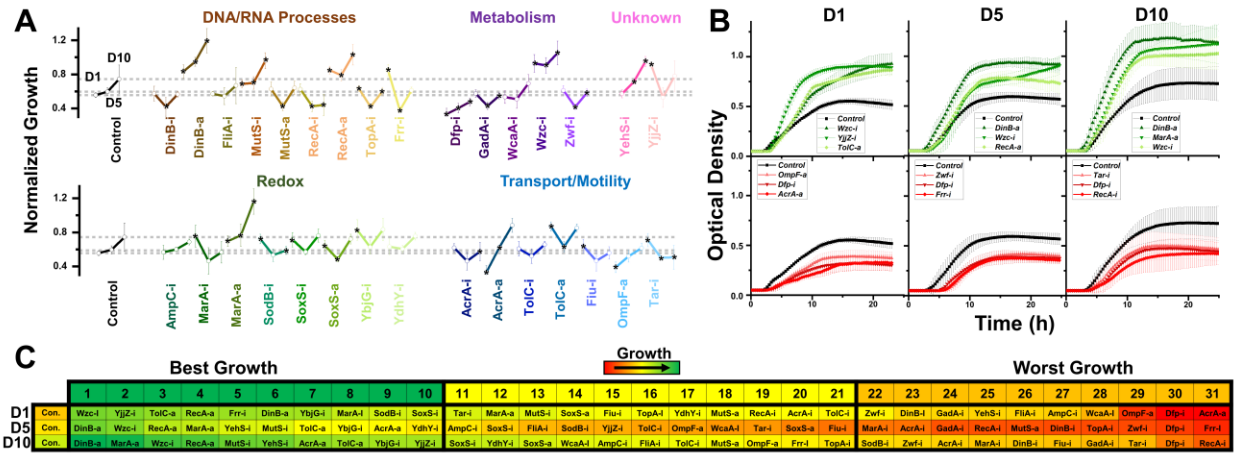


Figure 4.3 Normalized growth (maximum OD/starting OD) of *E. coli* harboring CRISPR gene perturbations during 0.5% vol/vol n-butanol exposure

(**A**) Change in growth of each strain over ten days of exposure, with quantification on days one (D1), five (D5), and ten (D10). Strains are organized based on pathways affected by perturbation. Dashed lines extend from the control for each experimental day. A two-tailed type II t-test was used to calculate significance ($P<0.05$) relative to the control on the same experimental day. (**B**) Growth curves of the three strains growing to the highest levels (green, top) or lowest levels (red, bottom) on D1, D5, and D10. (**C**) Organized rankings of strains with highest growth reached on each day, with the color scale to indicate relative growth. The top ten and bottom ten are indicated as “best growth” and “worst growth” respectively. All error bars represent the standard deviation of eight biological replicates.

The most prominent growth impacts were observed from metabolic pathway perturbations. Strains Dfp-i, Zwf-i, and GadA-i always fell within the ten worst performing strains. Each of these genes is involved in central metabolic pathways – *dfp* is essential for coenzyme A synthesis, *zwf* expresses the first enzyme of the pentose phosphate pathway, and *gadA* helps regulates glutamate levels. Disruption of such central metabolic pathways appears to be deleterious to butanol tolerance across all time points, and

should, therefore, be avoided. Conversely, inhibition of *wzc* exhibited the highest growth on day one and was always one of the top three strains throughout the experiment. *Wzc* is involved in colanic acid biosynthesis[215], and these results suggest that diverting metabolic flux from colanic acid could improve growth in n-butanol. Growth curves for this strain demonstrated significant improvement over the control on all days of the experiment (Figure 3B). Interestingly, we observed that Wzc-i also demonstrated a ~9-fold increase in chromosomal mutation rates, which could explain its consistent improvement in n-butanol tolerance (Figure S1).

Intriguing time-dependent impacts on growth were observed from perturbations of transport and motility genes. Strain AcrA-a exhibited detrimental growth on day one, but improved growth later in the experiment, while strain AcrA-i. exhibited reduced growth only on day ten. TolC-a and Fiu-i showed improved growth at the beginning of the experiment but had no significant impact by day ten. Tar -i helped growth on day one but actually resulted in lower growth later in the experiment. Activation of *ompF* also resulted in diminished growth in the beginning, but this could be explained by the aforementioned diminished growth in the absence of n-butanol. The complex impacts of perturbations of transport and motility genes could explain why previous attempts to improve *E. coli* n-butanol tolerance by heterologously expressed efflux pumps have failed[216], as the fitness impact of these genes appears to depend on time.

Perturbations of DNA and RNA processes. Very strong growth improvements were observed by activation of *dinB* and *marA*, while inhibition of these genes decreased growth. Perturbations of *mutS* exhibited the opposite effect, with inhibition improving growth and activation decreasing growth. TopA-i and Frr-i exhibited time-dependent

phenotype switching as many of the transport gene perturbations did, both slightly improving growth in the beginning while resulting in diminished growth at later time points. The complex phenotypic responses of these perturbation strains highlight the transitory impacts of CRISPR perturbations on overall growth and the need to optimize expression in a time-sensitive manner.

Relative to other perturbations, those impacting redox pathways exhibited less of an impact on growth phenotypes. The exception to this is activation of *marA*, which grew to the fourth and second highest OD on the days five and ten. Interestingly inhibition of *marA* also improved growth on day one. *MarA* has been demonstrated to be significantly upregulated during n-butanol exposure[134], so while results pertaining to MarA-a are as expected, MarA-i improved growth suggests more fine-tuning of the Mar regulon (particularly downstream genes) could result in even further n-butanol tolerance gains.

Finally, perturbation of the uncharacterized genes *yehS* and *yjjZ* point to the untapped potential for improved n-butanol tolerance and need for further investigation. Our previous work analyzing the transcriptome of *E. coli* adapted to n-butanol revealed that *yjjZ*, an uncharacterized gene suggested to express a small RNA, was significantly downregulated during exposure to n-butanol [17]. In accordance with this, strain YjjZ-i grew to the second highest levels on day one. However, this perturbation appears to have provided no benefit in the later part of the experiment. Conversely, strain YehS-i exhibited improved growth on days five and ten.

Overall, we noted that all strains adapted to n-butanol exposure over time; the average maximum ODs increased (0.65 ± 0.15 and 0.73 ± 0.19 on day one and ten respectively). The impact of gene expression perturbations appears to be significantly

time-sensitive, with many exhibiting benefits only in the short-term that were lost in the long-term. Supporting this observation is the fact that the control strain became one of the best strains over time; the control strain grew to the fifth lowest levels on day one but thirteen highest on day ten. This could suggest that CRISPR gene expression perturbations slightly impeded the strains' adaptive potential in the long-term. This is supported by previous work that has noted how epigenetic epistatic interactions might constrain adaptation [19,217,218], and could imply a tendency for perturbations to be detrimental to improving butanol tolerance in the long-term.

4.3.3 Impact of CRISPR Perturbations on E. coli Lag Times and Growth Rates During n-butanol Exposure

We also characterized the perturbations' impacts on lag times and growth rates on day one (Figure 4A), day five (Figure 4B) and day ten (Figure 4C). These results could point to interesting differences in growth between the perturbation strains upon exposure to n-butanol stress.

Interestingly, while growth was generally improved by perturbations relative to the control on day one, the opposite was true for lag times, which were generally extended: Twenty six strains exhibited significant increases in lag times (Figure 4A). Across the entire experiment, only four strains (Zwf-i, TopA-i, Tar-i, and Frr-i) consistently demonstrated increased lag times (Figure 4B, C). The best performing strain, DinB-a, began to exhibit decreased lag times on later time points.

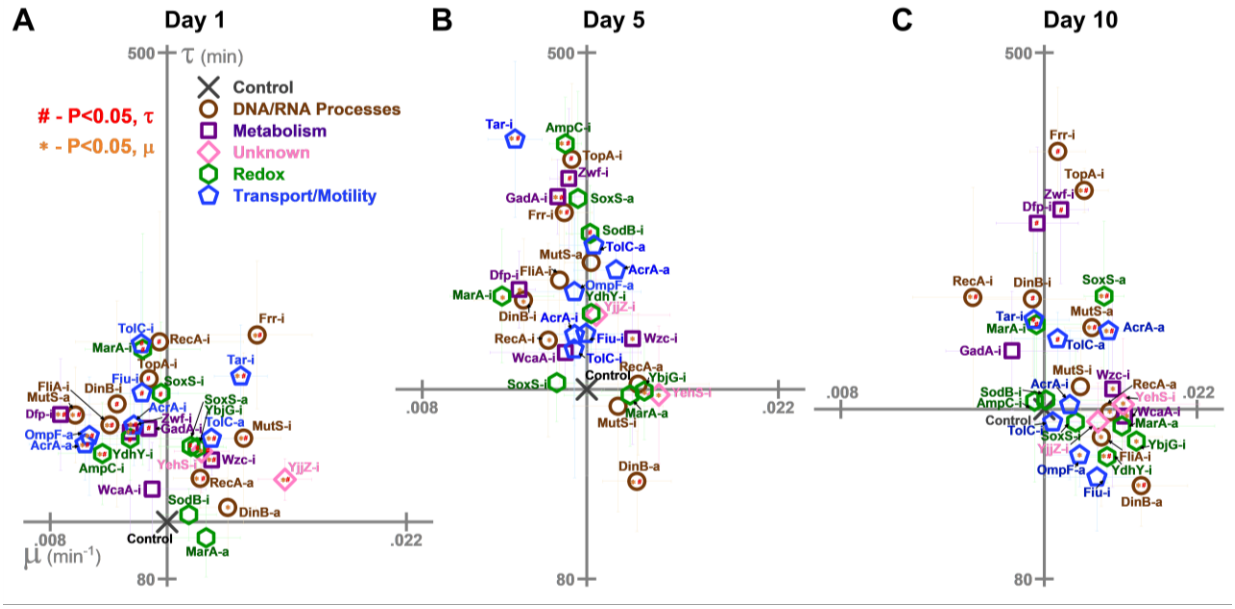


Figure 4.4 Growth rates (μ) and lag times (τ) of *E. coli* harboring CRISPR gene perturbations during 0.5% vol/vol n-butanol exposure

These growth characteristics were quantified on (**A**) day one, (**B**) day five, (**C**) and day ten of the experiment. Scales are set to intersect the control in each graph. A two-tailed type II t-test was used to calculate significance (P<0.05) relative to the control in growth rates (orange *) and lag times (red #). Error bars represent the standard deviation of eight biological replicates.

Many of the strains growing to the lowest levels also grew the slowest. Six of the ten worst growing strains on day one (Dfp-i, OmpF-a, AcrA-a, FliA-i, AmpC-i, and Zwf-i), six of the ten worst growing strains on day five (GadA-i, Frr-i, Dfp-i, MarA-i, DinB-i, and RecA-i), and the worst growing strain (RecA-i) on day ten all grew significantly slower on their respective days. Conversely, many of the best growing strains also grew the fastest. This includes six of the best ten growing strains on day one (YjjZ-i, Frr-i, TolC-a, Wzc-i, RecA-a, and DinB-a), four on day five (YehS-i, DinB-a, YbjG-i, and Wzc-i), and six on day ten (DinB-a, YbjG-i, RecA-a, YehS-i, Wzc-i, and AcrA-a).

Of note is the impact of gene expression activations over time. While lag times were relatively similar between both inhibition and activation constructs, gene activations

appeared to grow particularly faster over time. On day one, three activations improved growth rates while another three decreased growth rates. No gene activation slowed growth on days five and ten, while all gene activations aside from MarA-a and TolC-a significantly increased growth rates on day ten. These results suggest that better butanol tolerance can be achieved by waiting to activate gene expression once bacteria have adjusted to butanol exposure. This supports our earlier described butanol tolerance time-sensitivity to gene perturbations.

4.3.4 Impact of CRISPR Perturbations on E. coli Growth During n-hexane Exposure

One of the strong benefits of CRISPR gene perturbations is that it is relatively easy to test under diverse conditions, as everything is expressed from stable plasmids. We therefore performed the same growth assays of our CRISPR strains during exposure to a different biofuel, n-hexane, to demonstrate the power of this approach to identify gene targets under diverse conditions. We again quantified growth during the first, fifth, and tenth day of exposure to 10% vol/vol n-hexane (Figure 5). Due to the high-volatility of n-hexane, OD measurements were obscured during the initial few hours of growth, rendering lag time and growth rate calculations unreliable. Maximum ODs were still able to be measured in later hours of the experiment, allowing for determination of normalized growths (Figure 5A).

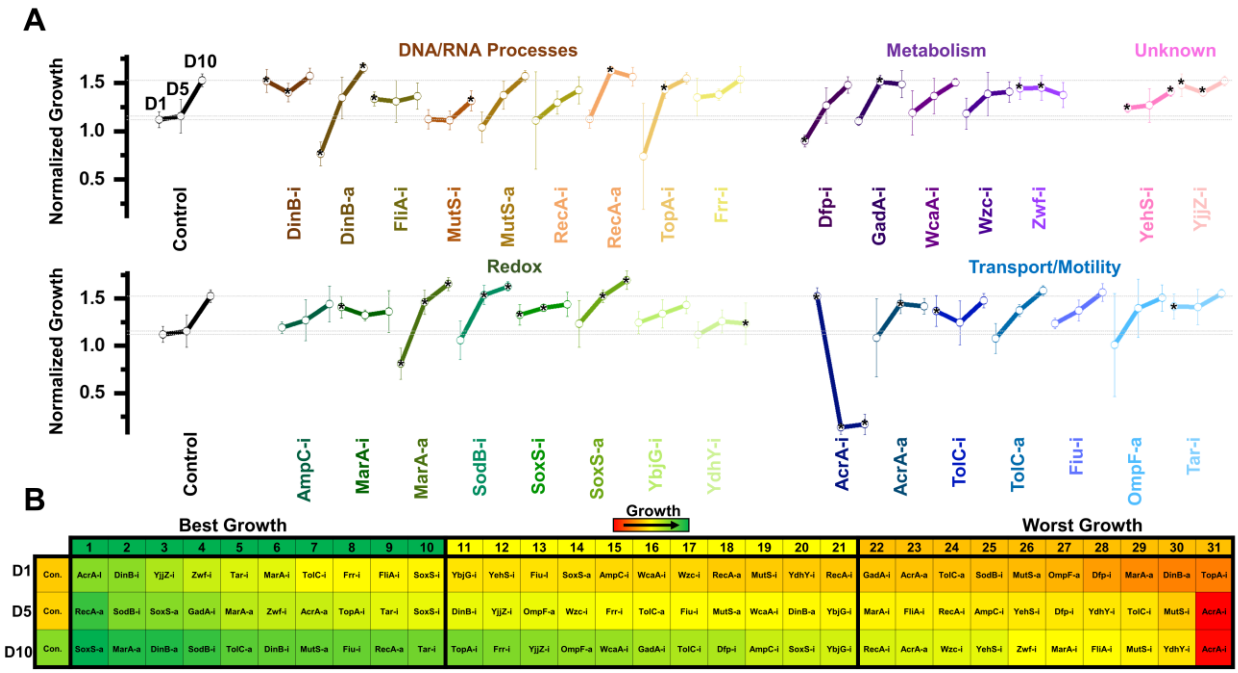


Figure 4.5 Normalized growth (maximum OD/starting OD) of *E. coli* harboring CRISPR gene perturbations during 10.0% vol/vol n-hexane exposure

(**A**) Change in the growth of each strain over ten days of exposure, with quantification on days one (D1), five (D5), and ten (D10). Strains are organized based on pathways affected by perturbation. Dashed lines extend from the control for each experimental day. A two-tailed type II t-test was used to calculate significance ($P<0.05$) relative to the control on the same experimental day. Error bars represent the standard deviation of four biological replicates. (**B**) Organized rankings of strains with highest growth reached on each day, with the color scale to indicate relative growth. The top ten and bottom ten are indicated as “best growth” and “worst growth” respectively.

Perturbations related to nucleic acid processes resulted in diverse responses across time points. This is most aptly demonstrated by strains DinB-i and DinB-a, which were the 2nd best and 2nd worst growers on day one respectively. DinB-i growth stayed virtually constant over the experiment. Conversely, DinB-a growth steadily improved, and eventually exhibited the third highest growth. Strains FliA-i, RecA-a, and TopA-i also showed transitory improvements in growth that only emerged on days one or five. The impact of *mutS* inhibition also appeared to be time-sensitive, having little impact on the

first day but growing to the 2nd and 3rd lowest levels on day five and ten respectively. Collectively, controlling the expression of nucleic acid processes appears to be highly time sensitive, making them ideal candidates for integration into inducible genetic circuits for improving n-hexane tolerance.

Metabolic perturbations resulted in a less pronounced impact on growth in n-hexane than they showed in n-butanol. Only inhibition of *dfp* on day one exhibited diminished growth but was able to recover on latter days. Inhibition of *gadA* and *zwf* appeared to provide a short-term benefit, but this too was lost by the final day of the experiment. This suggests that manipulation of metabolic pathways has less potential for optimization of n-hexane tolerance.

Conversely, perturbation of redox pathways elicited greater growth changes in n-hexane than was observed in n-butanol. Inhibition of both *marA* and *soxS* improved growth on day one, while activation of *marA* resulted in the third-lowest growth. This trend was notably reversed on days five and ten, with activation of these genes significantly improving growth. Indeed, these two strains were the best growing strains by the final day of the experiment, with the inhibited strains improving very little over time. Another interesting result is the inhibition of *sodB*, which grew to the 2nd and 4th highest levels on day five and ten respectively. A potential explanation for this phenomenon could be the degradation of n-hexane into n-hexanol in *E. coli* related to oxide levels, catalyzed by *sodB*'s gene product - superoxide dismutase. Alcohols are typically more toxic than alkanes due to their higher polarity[185], and reduced *sodB* expression could disrupt conversion into this more toxic chemical. Taken together, the responsiveness of *E. coli* to

redox perturbations in n-hexane suggests a potential area of focus for improving n-hexane tolerance.

Most perturbations of transport-related genes had little impact on growth, with the prominent exception of *acrA* inhibition. This strain exhibited the highest optical densities after the first day of n-hexane exposure, but every replicate died by the fifth day of the experiment (Figure 5B). The AcrAB-TolC efflux pump is known to export solvents such as hexanes from inside *E. coli*[219]. The eventual death caused by *acrA* inhibition demonstrates that engineered transcriptome changes are sufficient to mimic total gene knockout phenotypes, suggesting that CRISPR perturbations could be used instead of genomic changes. The established connection between this efflux pump and n-hexane tolerance also explains the higher ODs upon activation of *acrA* on day five. However, this would run counter to results of *tolC* perturbation whose activation never significantly impacted ODs, and whose inhibition also increased ODs on day one. These results suggest a more complicated underlying story, and that temporal manipulation of the AcrAB-TolC efflux pump could offer an interesting strategy to improve bacterial tolerance to alkanes.

Finally, inhibition of the uncharacterized genes *yehS* and *yjjZ* during exposure to n-hexane resulted in similar phenotypes as observed in n-butanol – both perturbations improved growth on day one, with *yjjZ* demonstrating the third highest growth. As perturbation of these genes was again able to improve biofuel tolerance, our data indicate that these genes are highly promising candidates for future research.

4.3.5 Gene Knockout Phenotypes Corroborate CRISPR Perturbation Results

To corroborate our CRISPR perturbation results, we examined the growth of fifteen gene knockouts in the presence of (Figure 6A) 0.5% n-butanol and (Figure 6C) 10.0% n-hexane exposure. This included three genes related to DNA/RNA processes (*dinB*, *mutS*, and *recA*), the metabolism gene *wzc*, the unknown genes *yehS* and *yjjZ*, five redox-related genes (*marA*, *sodB*, *soxS*, *ybjG*, and *ydhY*), and four transport or motility-related genes (*acrA*, *tolC*, *fiu*, and *tar*).

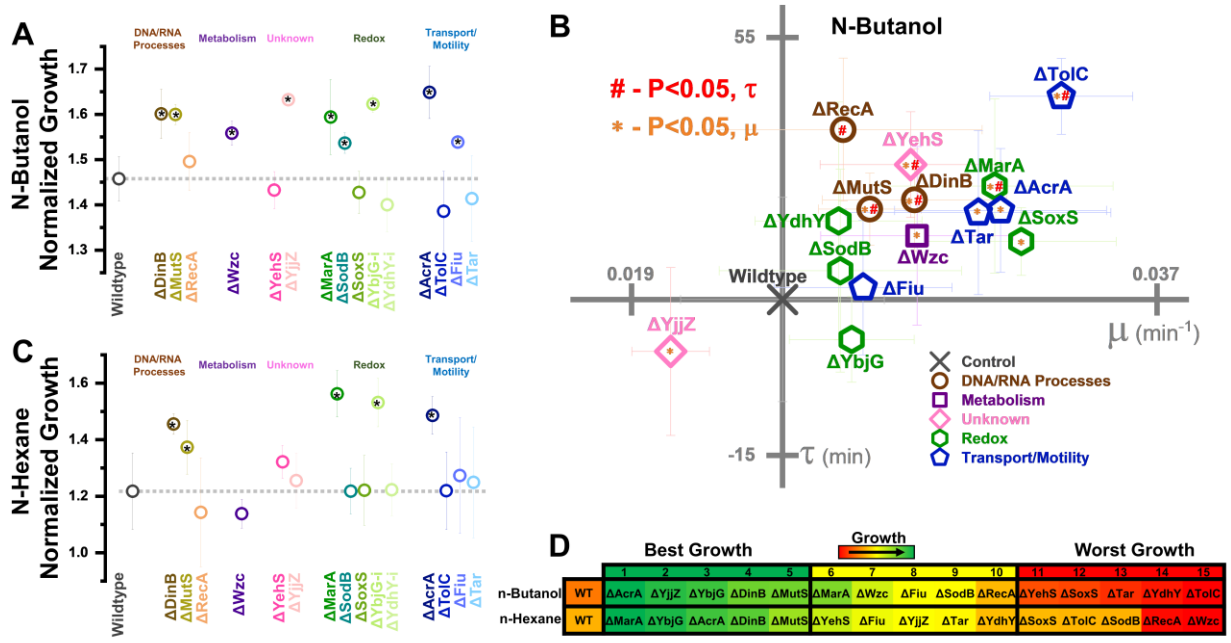


Figure 4.6 Growth of gene knockouts in relation to wildtype *E. coli* BW25113

Normalized growth (maximum OD/starting OD) of knockouts in (**A**) 0.5% vol/vol n-butanol or (**C**) n-hexane. Strains are organized based on pathways affected by perturbation. Dashed lines extend from the control A two-tailed type II t-test was used to calculate significance (P<0.05) relative to the control. (**B**) Growth rates (μ) and lag times (τ) of knockouts during 0.5% vol/vol n-butanol exposure. Axes are set to intersect the control in each graph. A two-tailed type II t-test was used to calculate significance (P<0.05) relative to the control in growth rates (orange *) and lag times (red #). (**D**) Organized rankings of strains with highest growth reached on each day, with the color scale to indicate relative

growth. The top five and bottom five are indicated as “best growth” and “worst growth” respectively. All error bars represent the standard deviation of four biological replicates.

Nine knockouts exhibited significant increases in growth over the wildtype during n-butanol exposure. Of these, seven (ΔMutS, ΔWzc, ΔYjjZ, ΔSodB, ΔYbjG, ΔAcrA, and ΔFiu) exhibited similar phenotypes as the corresponding CRISPR perturbations. While ΔDinB improved growth, activation of DinB gene expression also resulted in improved growth. The perturbation results are consistent with previous work showing that DinB overexpression improved long-term adaptive potential towards n-butanol[220]. That DinB knockout also improves growth indicates that short term butanol tolerance may be more nuanced and require further study.

In a similar vein, ΔMarA resulted in improved growth. This corroborates our results showing improved n-butanol growth caused by *marA* inhibition, but runs counter to results showing its activation also improving growth. The rapid over-expression of MarA immediately after n-butanol exposure has been previously reported, supporting our MarA-a results[134]. However, it has also been demonstrated that total knockout of MarA’s repressor, MarR, resulted in diminished growth in n-butanol[221]. This appears to corroborate our results showing increased growth of ΔMarA and MarA-i. A potential explanation for these conflicting methods of improved n-butanol tolerance could stem from the fact that the *marRAB* operon is known to exhibit stochastic pulsing behavior[114]. Further synthetic tuning of the mar regulon may therefore provide for greater n-butanol tolerance.

Interestingly, ΔYjjZ actually exhibited the slowest growth rate in n-butanol (Figure 6B). This runs contrary to the CRISPR perturbation results, where its inhibition resulted

in the fastest growth rate of all strains on day one. Of the remaining gene knockouts, nine exhibited significant increases in growth rates relative to the control, five of which (ΔAcrA, ΔMarA, ΔWzc, ΔDinB, and ΔMutS) also exhibited increased growth over the control. Only three of these strains (MutS-i, Tar-i, and YehS-i) actually exhibited an increased growth rate in the CRISPR perturbation context. The remaining four strains (ΔYehS, ΔTolC ΔSoxS ΔTar) were four of the five worst growing strains (Figure 6D).

Six strains exhibited increased lag times in n-butanol (ΔRecA, ΔMutS, ΔYehS**,** DinB, ΔMarA, and ΔTolC), and no strain exhibited decreased lag time. While CRISPR inhibitions of each of these strains also exhibited increased lag times on day one, it should be noted that most CRISPR perturbation strains increased lag times over the control. This trend was broadly recapitulated in gene knockouts, suggesting that an underlying phenomenon is indeed causing these genetic manipulations to increase lag times during n-butanol exposure.

Growth in n-hexane was improved by five knockouts (ΔDinB, ΔMutS, ΔMarA, ΔYbjG, and ΔAcrA), of which all but MutS-i and YbjG-i improved growth in the CRISPR perturbation context (Figure 6C). The strongest improvements in n-hexane tolerance were again related to redox-related genes, as ΔMarA and ΔYbjG were the top two growing strains (Figure 6D). The large improvements in growth observed from SodB-i only emerged in later time points and could be why ΔSodB showed no significant differences from the control.

4.3.6 CRISPR Perturbations Retain Growth Impacts Despite a Hyper-mutator Phenotype

One frequent criticism of CRISPR perturbation strategies is the potential for mutations to arise that inactivate the system. As bacteria are continually exposed to stressful conditions, they inevitably accumulate mutations; a mutation in the CRISPR expression system, such as a deletion in the sgRNA, could deactivate the perturbation. This is especially concerning if the perturbation is detrimental at any point during growth, which we have demonstrated is frequently the case. To address these concerns, we designed a system that biases the sgRNA plasmid towards hyper-mutation rates to illustrate how mutation rates can affect the efficacy of CRISPR perturbation strategies.

We accomplished this by incorporating an error-prone version of Polymerase 1 (Pol1) with greatly diminished fidelity into our CRISPR perturbation strains on the plasmid expressing dCas9 (Figure 7A, see Methods). We transformed each CRISPR construct into a strain of E. coli with temperature-sensitive wild-type Pol1 that fails to express at temperatures above 30°C, causing this error-prone version of Pol1 to overtake its functionality. Pol1 initiations replication of ColE1 plasmids, while having no role in replicating plasmids using the pSC101 ori that drives dCas9 and dCas9-ω expression. Thus, the sgRNA plasmid is significantly more prone to accumulating mutations in this system. Previous work designing this error-prone Pol1 estimated that in vivo mutation rates are increased ~80,000-fold above basal levels for at least 3 kb beyond the ColE1 ori (Camps et al., 2003), with only 3-5 fold increases in mutation rates of the chromosome at large. We confirmed that integration of the error-prone Pol1 into our bacteria increased chromosomal mutation rates ~3-fold (Figure 7B), in line with these published results.

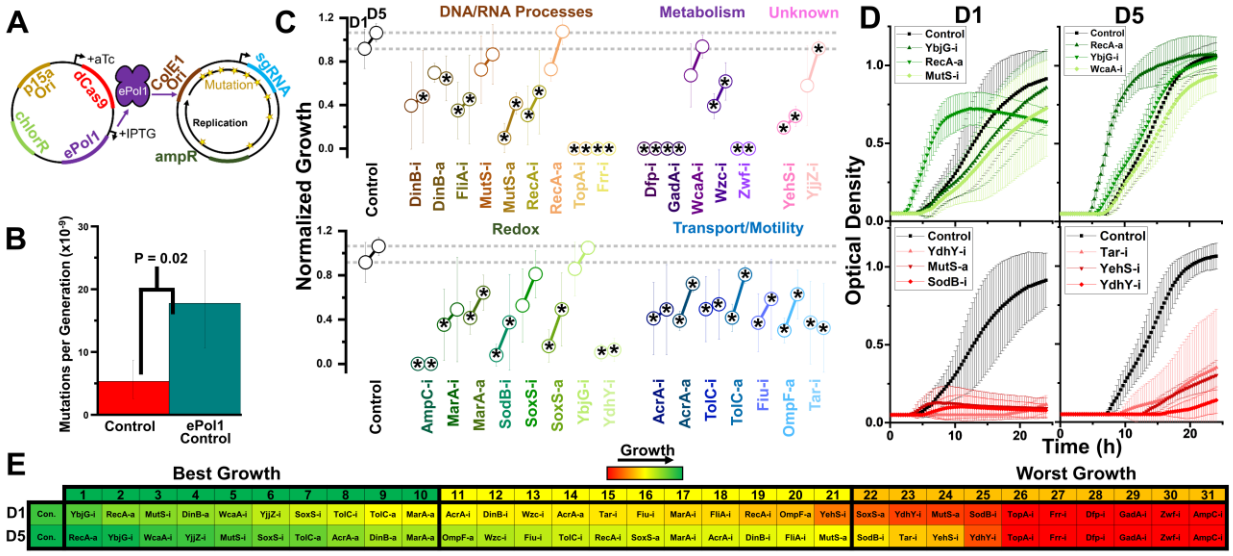


Figure 4.7 Design of a hyper-mutator strain of *E. coli* for targeted error-prone replication of the sgRNA plasmid, and subsequent growth of these strains in 1.0% vol/vol n-butanol exposure

(**A**) We move dCas9 and dCas9-ω onto a plasmid expressing IPTG inducible error-prone Pol1 in a strain of *E. coli* expressing temperature-sensitive native Pol1. During growth at 37 °C, error-prone Pol1 is expressed, causing low fidelity replication of plasmids with the ColE1 ori. This imparts significant mutations of the sgRNA plasmids with minimal impact on the dCas9/dCas9-ω plasmid or genome at large. (**B**) Whole-genome mutation rates of the control strain and the hyper-mutator control strain. Error bars represent the standard deviation of 32 technical replicates. A two-tailed type II t-test was used to calculate the statistical difference between the strains. (**C**) Normalized growth (maximum OD/starting OD) of hyper-mutator *E. coli* harboring CRISPR gene perturbations during 1.0% vol/vol n-butanol exposure. Change in the growth of each strain over five days of exposure, with quantification on days one (D1) and five (D5). Strains are organized based on pathways affected by perturbation. Dashed lines extend from the control for each experimental day. A two-tailed type II t-test was used to calculate significance ($P<0.05$) relative to the control on the same experimental day. Error bars represent the standard deviation of four biological replicates. (**D**) Growth curves of the three strains growing to the highest levels (green, top) and lowest levels (red, bottom) on D1 and D5. (**E**) Organized rankings of strains with the highest growth reached on each day, with the color scale to indicate relative growth. The top ten and bottom ten are indicated as “best growth” and “worst growth” respectively.

By incorporating error-prone Pol1 into our CRISPR perturbation system, we could simulate how prolonged mutation might impact the efficacy of our perturbations towards

engineering biofuel tolerance. We implemented this system into each of our CRISPR perturbation strains, and again tested their impacts on growth during biofuel exposure. We focused on n-butanol stress due to the aforementioned difficulty of quantifying growth rates and lag times in n-hexane. n-Butanol was doubled to 1% vol/vol to increase the selective pressure driving mutations. As the majority of impacts in n-butanol emerged by day five of the experiment, we limited the experiment to five days of exposure, quantifying growth phenotypes on the first and last day (Figure 7C and Figure S2). Finally, we included ampicillin selection to ensure that the sgRNA plasmids were not lost completely due to mutations, thereby biasing mutations solely towards the portion of the plasmid responsible for expressing the sgRNA.

Of immediate note is the failure of six strains to grow at this higher concentration of n-butanol. That these strains were unable to recover by day five suggests that even the hyper-mutation rates of this system were not sufficient to recover the detrimental phenotype, and suggest short-term stability of the CRISPR perturbation system even when causing reduced fitness. These strains harbored inhibition constructs targeting the genes ampC, gadA, dfp, zwf, topA, and frr. Each strain was one of the ten worst growing strains on day one (Dfp-i, AmpC-i, GadA-i, and Zwf-i), day five (Dfp-i, Zwf-i, TopA-i, Frr-i, and GadA-i), or day ten (Dfp-i, GadA-i, and Zwf-i) at lower n-butanol concentrations in the absence of the hyper-mutator phenotype, suggesting that this is a result of doubling butanol levels. Half of these genes were metabolism-related (dfp, gadA, and zwf). Two of these other genes, topA and frr, are essential for growth, and their inhibition likely synergized with the toxic effects of butanol to induce cell death. The final gene, ampC, is an inherent periplasmic beta-lactamase, and to our knowledge has never been linked to

biofuel tolerance. The death of these six strains suggests that expression of these genes should strongly be considered when optimizing n-butanol tolerance.

The control strain was the top performing strain on day one and second best on day two, indicating perturbations were largely detrimental at this increased n-butanol concentration. We plotted the growth curves of the top and bottom three perturbation strains on each day, excluding the six strains that died during growth (Figure 7D). RecA-a growth curves strikingly show a significantly faster growth rate and reduced lag time (confirmed in Figure S2), despite not reaching higher concentrations at the end of each day.

Overall, most detrimental perturbations did not reach control level growth after five days of exposure (Figure 7E). This suggests that despite the hypermutator phenotype, detrimental phenotypes remained detrimental. To ensure that this was not a result of the failure of the hyper-mutator phenotype, we sequenced sixteen individual colonies of the moderately detrimental OmpF-a perturbation. We observed four mutations in the sgRNA plasmid, none of which were located in the actual sgRNA coding sequence. This led us to estimate a mutation rate of $8.36 * 10^{-6}$ mutations per nucleotide per generation, or a ~2600 fold increase in mutation rates above basal levels (see Methods). While this estimate is significantly lower than the 80,000 fold increase reported for the error-prone Pol1 system, we can confidently report a large increase in mutation rates. That no mutations arose inactivating the perturbation after five days suggests that spontaneous mutations inactivating the system are less likely than might be initially predicted. Collectively, this data demonstrates that CRISPR perturbations are stable even in a hyper-mutator strain.

4.4 Discussion

This study applies recent advances in synthetic biology to harness the untapped potential of altering gene expression states in biofuel applications. We explored 31 unique CRISPR inhibitions and activations of a diverse set of bacterial genes and quantified their impacts on *E. coli* growth during exposure to two common biofuels, n-butanol and n-hexane. We identified a number of strong gene candidates whose expression could be engineered to enhance biofuel tolerance such as RecA-a, YjjZ-i, and Wzc-i.

A number of these perturbations' growth impacts were time-sensitive, suggesting that they could be implemented into temporal gene circuits to improve biofuel production capacity. This possibility is gaining popularity due to the relatively facile ability to integrate CRISPR perturbations into such circuits[55,222]. Indeed, as efforts to improve biofuel tolerance have stalled, the need for genetic circuits to manipulate transcription at particular time points has been recognized yet relatively unexplored[223]. CRISPR perturbation can make such genetic circuits attainable, and this study presents the first evidence towards this goal.

Engineering gene expression has long been a goal for biotechnology application. However, previous approaches for accomplishing this including manipulation of promoter elements[56] or riboswitches[224] have each suffered from their own unique drawbacks that have made them difficult to implement in practice. Perhaps the most notable limitation is the reliance upon stable alteration of genomes. CRISPR perturbations, on the other hand, can be implemented without direct manipulation of the bacteria's genome via plasmids or extracellular delivery of the CRISPR machinery. This can also be much easier to engineer

in practice than direct mutations of the genome, which has frequently proven difficult in a number of promising biofuel producing microorganisms[205].

Furthermore, we have shown here that simple knockdowns can be sufficient to impart significant growth phenotypes that mimic total gene removal. Finally, multiplexing sgRNAs to target one gene multiple times or multiple genes at the same time is exceedingly simplified by the introduction of numerous unique sgRNAs simultaneously and is gaining significant attention[64,225]. Combining the best perturbations presented in this study could conceivably be done to raise tolerance levels even further.

We also demonstrated that despite artificially amplifying sgRNA mutation rates ~2600-fold, CRISPR perturbations induced similar growth phenotypes. This suggests that CRISPR perturbations can be maintained stably for prolonged periods without loss of functionality. As such perturbations begin to be applied towards biotechnology purposes, such long-term stability will be essential to maintain the desired phenotype. Our data suggest that loss of sgRNA functionality, even if detrimental, is unlikely in the short to medium term.

Going forward, we envision that this hyper-mutation system could be employed towards the directed evolution of novel sgRNA targets, improving fitness without requiring *a priori* knowledge. Over long periods, detrimental mutations to the sgRNA would be selected against, while the rare beneficial mutations that redirect CRISPR perturbations to new targets would be selected for in a manner highly analogous to traditional directed evolution approaches[226]. Various alterations could be made to our hyper-mutator system to make this approach more viable. For instance, It has been reported that mutagenesis from error-prone Pol1 is strongest during stationary phase, and suggested that mutations

are concentrated in locations closest to the origin of replication (~700 bp)[210,211]. Removal of extraneous DNA segments would increase the likelihood of targeted mutations towards the 20nt target sequence of the sgRNA. It may be beneficial to express the sgRNA in its native, two component fashion where tracrRNA is expressed separately from the target sequence: expressing the tracrRNA on a separate plasmid would ensure its structure is not lost by mutation. Growth in a bioreactor to maintain steady-state conditions would ensure maximum mutation rate, and would likely impart a more consistent selective pressure to obtain beneficial mutations. This would also allow for a controlled increase of butanol concentration, as the constant butanol concentration used in this study likely limited further selection.

4.5 Materials

4.5.1 CRISPR Plasmid and Strain Construction

Addgene plasmids #44249 and #44251 were used for expression of dCas9 and sgRNA respectively. Native 44251 targets the ORF of RFP, which is not present in any of the strains used in this study and was therefore used as the nonsense control sgRNA target sequence. Plasmid pPO-dCas9ω was constructed in a previous study[19] and used for expression of dCas9-ω alongside 44251. Unique sgRNA targets were constructed by PCR amplifying cloning inserts replacing the RFP target sequence with the new target sequence for each gene. Inserts were flanked with SpeI and ApaI restriction sites. Plasmid 44251 digested with SpeI and ApaI (New England Biolabs) was used as the cloning backbone. Digested inserts were ligated alongside this backbone and transformed into electrocompetent NEB 10-β. Final constructs were recovered using

Zyppy Plasmid Miniprep Kit (Zymo Research Corporation) and confirmed by sequencing before transformation into chemically competent *E. coli* MG1655 (ATCC 700926) harboring either dCas9 or dCas9-ω plasmids for gene repression or activation respectively. Exact gene targets for each sgRNA are listed in Table S1 (Supporting Information). The successful perturbation of gene expression using this CRISPR system was confirmed using quantitative real-time PCR in previous studies[16,17,19].

4.5.2 Error-prone Strain Construction

Strain JS200 expressing temperature-sensitive polA was obtained from Addgene (#11722) harboring the pEP Pol1 plasmid (error-prone polA D424A, I709N, A759R with reduced fidelity) with chloramphenicol resistance marker. The plasmid was miniprepped from the strain, after which the strain's plasmid was removed by growing for five days at 30 °C in 3 mL LB cultures, with 1:1000 dilution into fresh culture every 24 hours. The culture was streaked on plain LB agar plates at the end of this exposure period to obtain individual colonies. These colonies were screened for successful plasmid removal by plating in both the presence and absence of chloramphenicol. A colony that grew only in the absence of chloramphenicol was picked and saved to obtain strain JS200 with no plasmid.

Plasmids dCas9 and dCas9-ω were PCR amplified as Gibson Assembly inserts, while plasmid pEP Pol1 was PCR amplified as a Gibson Assembly backbone. Primers are listed in Table S2 (Supporting Information). Successful PCR products were gel extracted, and Gibson Assembly was performed to insert pEP into dCas9 and dCa9-ω plasmids. Gibson controls using only insert or backbone were run in parallel to confirm

successful assembly. Constructs were transformed into electrocompetent NEB10-β, plasmids were recovered and run on a gel to confirm appropriate sizes, and submitted for sequencing confirmation. Plasmids were then transformed into empty chemically competent JS200, with overnight growth at 30 °C with cm selection. Successful transformants of Pol1-dCas9 and Pol1-dCas9-ω were picked and grown overnight at 30 °C. Each strain was made chemically competent and immediately transformed with each of the individual sgRNA targets, with growth at 30 °C. To prevent excessive mutation before the start of the experiment, transformation plates were used directly to inoculate 4 biological replicates grown overnight at 30 °C for the experiment represented in Figure 7. Experiments using these strains included 100 µM IPTG to drive expression of error-prone Pol1.

4.5.3 Growth and Media Conditions

All cultures were grown in Lennox Luria-Bertani Broth (LB) (Sigma-Aldrich). Media was supplemented with ampicillin (amp, 100 µg/mL) to maintain a selection of sgRNA plasmids, or supplemented with chloramphenicol (cm, 35 µg/mL) to maintain a selection of dCas9/dCas9-ω/pEP Pol1 plasmids. Unless noted, amp and cm were always included in media. Growth of gene knockout strains was performed without supplementation of any antibiotic. Expression of dCas9/dCas9-ω during experiments was driven by supplementation of 50 ng/mL aTc. Expression of error-prone Pol1 during experiments was driven by supplementation of 100µM IPTG. All cultures were grown at 37 °C, with shaking at 225 rpm unless otherwise noted. Growth at 30 °C was used during cloning of the error-prone strains in order to drive expression of wild-type Pol1.

4.5.4 Growth Assays During Biofuel Exposure

For all growth experiments, individual colonies of normal CRISPR-perturbation constructs or gene knockouts were inoculated into 100 µL cultures in 384 well flat-bottom microplates and grown overnight for 16 hours. Cultures were then diluted 1:100 into fresh 100 µL cultures supplemented with aTc (except for gene knockout strains) and grown for 24 hours. Cultures were then diluted 1:100 into fresh 100 µL cultures supplemented with either no biofuel (Figure 2), 0.5% vol/vol n-butanol (Figures 3, 4, and 6) or 10% vol/vol n-hexane (Figure 5) , and grown in a GENios plate reader (Tecan Group Ltd.) operating under Magellan software (version 7.2) with shaking every 16.6 min before OD measurement every 20 min. Cultures were grown for 24 hours, and data from the microplate run was used to determine growth characteristics on “day 1” of the experiment. The significant volatility of n-hexanes disrupted OD measurements during the first ~5 hours of growth, hence the exclusion of lag times and growth rates for n-hexane data. For CRISPR perturbation strains, after 24 hours of growth, cultures were diluted into fresh media and grown in a regular shaking incubator for days two-four and six-nine. Cultures were again grown in the plate reader on days five and ten of the experiment to capture changing growth characteristics over time.

For Figure 7, four individual colonies of CRISPR-perturbation constructs expressing error-prone Pol1 in JS200 were inoculated directly from transformation plates into 100 µL cultures supplemented with amp and cm and grown for 16 hours overnight. Cultures were diluted 1:100 into fresh 100 µL cultures supplemented with aTc, IPTG, and 1.0% vol/vol n-butanol (increased to exacerbate selective pressure) and grown for 24

hours in the plate reader for day one growth measurements. OD measurements were repeated in the microplate reader for day five of the experiment.

4.5.5 Mutation Fluctuation Assay

Whole-genome mutation rates were determined using the Luria-Delbruck method of identifying spontaneous rifampicin resistance. An individual colony of strains for this experiment was inoculated in 3 mL of LB and grown overnight for 16 h without ampicillin or chloramphenicol selection. Each culture was then normalized to the same OD and grown, and diluted 1:10,000 into 35 parallel 100 µL cultures supplemented with 50 ng/mL aTc and grown for another 24 h. Three cultures of each strain were used to determine colony forming units, revealing overall viable cells per strain. Of the remaining 32 cultures, 50 µL of each were plated on LB agar supplemented with 100 µg/mL rifampicin and grown for 24 h. Colonies were then calculated, and mutation rates were estimated using the FALCOR web tool [177].

4.5.6 Determination of sgRNA Mutation Rate via Sequencing

To quantify mutation rates of the sgRNA plasmids in the error-prone polymerase system, twelve JS200 *E. coli* cells harboring error-prone Pol1 alongside dCas9-ω and the *ompF* activation sgRNA were exposed to 1.0% n-butanol for five days using the protocol listed above. After five days of exposure, replicates were streaked on plain LB plates and grown overnight. Sixteen individual colonies were selected from one replicate showing the greatest amount of growth, grown overnight in 5 mL LB, and miniprepped to recover the sgRNA plasmids. These plasmids were submitted for standard Sanger sequencing

(GENEWIZ) using the primer 5'–aaataggcgtatcacgaggc– 3'. Sequencing results revealed ~900 nucleotides of reliable sequence per sample. Mutations were identified via BLAST alignment, revealing a total of four point mutations in all sixteen samples. From this data, it was estimated that four mutations per 900 * 16 nucleotides or a mutation profile of 2.78 * 10^{-4} mutations per nucleotide. As a 1:100 dilution of *E. coli* into fresh LB has been estimated to result in roughly ~6.64 new generations per day[227], we estimate that 33.2 generations of bacteria passed throughout the five-day evolution experiment. This gives an estimated mutation rate of 8.36 * 10^{-6} mutations per nucleotide per generation of the sgRNA plasmid. The established mutation rate of *E. coli* is 3.2 * 10^{-9} mutations per nucleotide per generation [175] (within error of our calculated mutation rate of the control in Figure 6B), suggesting that our system exhibited a ~2600 fold increase in mutation rate. While this level is clearly higher than basal levels, it is significantly lower than the reported ~80,000 fold increase[210]. This is likely due to a reduction in mutagenesis efficiency after reaching stationary phase, as has been reported[211]. Improved mutation rates could likely be achieved by maintaining cultures in exponential phase through growth in a bioreactor.

4.5.7 Growth Analysis

ODs were normalized to blank-LB cultures from the same day of each experiment and subsequently normalized to starting ODs. These values were used to determine lag times, growth rates, and maximum ODs using the program GrowthRates version 1.8[161].

4.5.8 Statistical Analysis

All P values reported were calculated using a standard two-tailed type II student's t-test in comparison to the RFP-targeting control strain within each graph, with a significance value of $\alpha = 0.05$. All normalized growth, optical density, growth rate, and lag time error bars represent standard deviations of four or eight biological replicates as indicated. Error bars of mutation fluctuation analysis represent standard deviations of 32 technical replicates.

4.6 Author Contributions

P.B.O. conceived of the study, performed all experiments and data analysis. P.B.O. and A.C. wrote and edited the paper.

4.7 Supplementary Information

4.7.1 Supplementary Tables

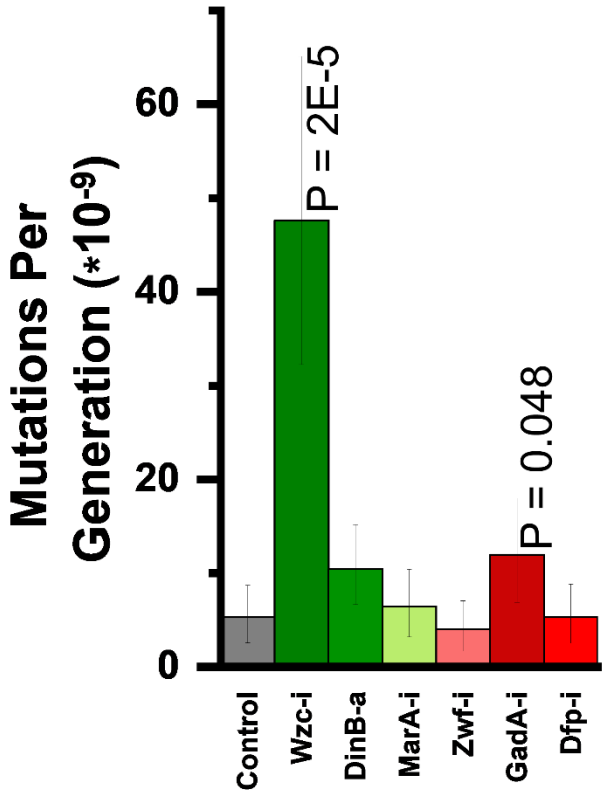
Supplementary Table 4.1 The unique 20 nt sgRNA target used to activate or inhibit gene expression

Gene Perturbation	Sequence
acrA-i	AGCATCAGAACGACCGCCAG
acrA-a	GAGCCACATCGAGGATGTGT
ampC-i	TAGGCGGGCCGGATTTACAT
dfp-i	GTGATAAAATCGCCAACTTC
dinB-i	ATTGTCGCGCATCTCCACTG
dinB-a	GCAAAAGCTGGATAAGCAGC
fiu-i	GTCCCTTTAACGCTAACAAA
fliA-i	TCGCTCACAAATAGGTAATG
frr-i	AGCCCTGATTAAACATATTA
gadA-i	AAAGTAGGATTTATCCGCAA
marA-i	CCAGTCCAAAATGCTATGAA
marA-a	GTTTTGTTCAATGCGATGCA
mutS-i	CTGCTGCATCATGGGCGTAT
mutS-a	GCAAGTACGCAAAATTGTAT
ompF-a	GTAACCAAAAGTAAAATTTA
recA-i	TACCAAATTGTTTCTCAATC
recA-a	CCGTGATGCGGTGCGTCGTC
sodB-i	ACCATATGCTAAAGATGCTC
soxS-i	CTACATCAATGTTAAGCGGC
soxS-a	GCGTTTCGCCACTTCGCCGG
tar-i	CGCGGATACGGTTAATCATA
tolC-i	GGCTCAGGCCGATAAGAATG
tolC-a	AGCAGTCATGTGTTAAATTG
topA-i	CTGGCAACGAGTTACCGATA
wcaA-i	TCTCAATCTATATGCCGACC
wzc-i	CAACATGCCGCTCCGGTAAC
ybjG-i	TATCTCTCTCTAAGTTTAAA
ydhY-i	GATCGTCCACTATTAGATAT
yehS-i	CGCACGCGATGTAAAACTTT
yjjZ-i	ATCATGTTGCAACGTACGCT
zwf-i	GTATACTTGTAATTTTCTTA
RFP (Control)	AACTTTCAGTTTAGCGGTCT

Supplementary Table 4.2 Cloning primers used for transferring the error-prone Pol1 sequence into dCas9/ dCas9-ω plasmids.

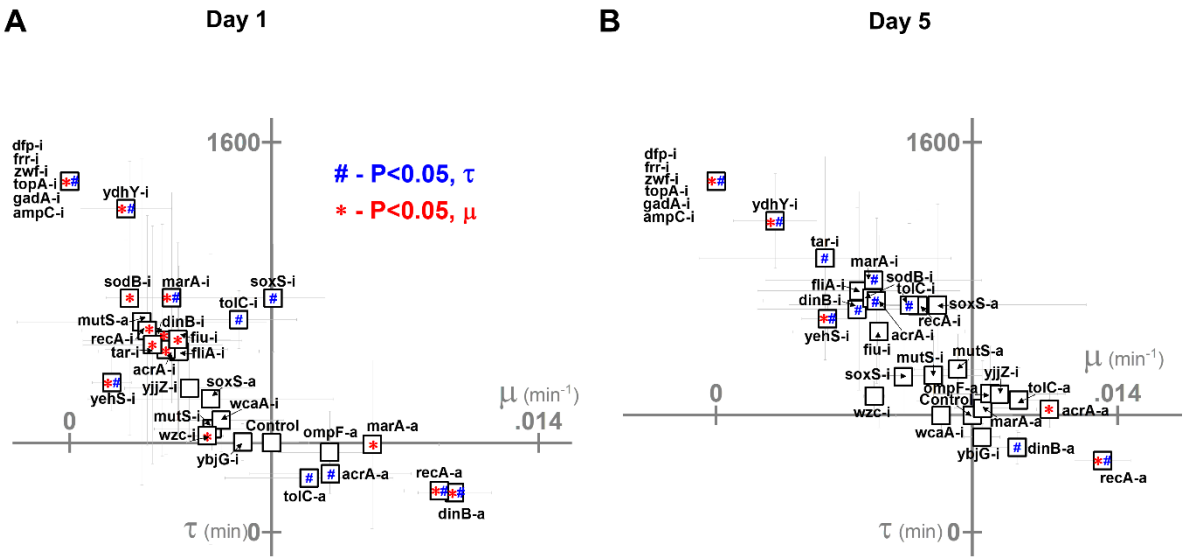
Primer	Sequence
dCas9/ω Fwd	AATAACCTAGGAATCATGGCAATTCTGGAAGAAATAGCGC
dCas9/ω Rev	AATTTGACGTCTTACATGCTGTTCATCTGTTACATTGTCG
ePol1 Fwd	ACAGATGAACAGCATGTAAGACGTCAAATTTTAAGACCCACTTTCACATT
ePol1 Rev	TTCCAGAATTGCCATGATTCCTAGGTTATTTCTAGTAGAGAGCGTTCACC

4.7.2 Supplementary Figures



Supplementary Figure 4.1 Mutation fluctuation assay

Mutation rates of three strains growing well in n-butanol (green) and poorly in n-butanol (red). Data was collected as in Figure 7B, using 32 biological replicates. A two-tailed type II t-test was used to calculate statistical differences between strains, and those that exhibited significant differences ($P<0.05$) are labeled with their corresponding P-value. Error bars indicate 95% confidence interval.



Supplementary Figure 4.2 n-butanol impact on growth

Growth rates (μ) and lag times (τ) of hyper-mutator *E. coli* harboring CRISPR gene perturbations during 1.0% vol/vol n-butanol exposure. These growth characteristics were quantified on (A) day one and (B) and day five of the experiment. Scales are set to intersect the control in each graph. A two-tailed type II t-test was used to calculate significance (P<0.05) relative to the control in growth rates (red *) and lag times (blue #). Error bars represent the standard deviation of four biological replicates.

Chapter 5: CRISPR perturbation of gene expression alters bacterial fitness under stress and reveals underlying epistatic constraints

Reprinted with permission from Otoupal, P. B., Erickson, K. E., Bordoy, A. E., and Chatterjee, A. (2016) CRISPR Perturbation of Gene Expression Alters Bacterial Fitness under Stress and Reveals Underlying Epistatic Constraints. ACS Synthetic Biology (vol 6), pg. 94-107 DOI: 10.1021/acssynbio.6b00050.

5.1 Abstract

The evolution of antibiotic resistance has engendered an impending global health crisis that necessitates a greater understanding of how resistance emerges. The impact of non-genetic factors and how they influence the evolution of resistance is a largely unexplored area of research. Here we present a novel application of CRISPR-Cas9 technology for investigating how gene expression governs the adaptive pathways available to bacteria during the evolution of resistance. We examine the impact of gene expression changes on bacterial adaptation by constructing a library of deactivated CRISPR-Cas9 synthetic devices to tune the expression of a set of stress-response genes in Escherichia coli. We show that artificially inducing perturbations in gene expression imparts significant synthetic control over fitness and growth during stress exposure. We present evidence that these impacts are reversible; strains with synthetically perturbed gene expression regained wild-type growth phenotypes upon stress removal, while maintaining divergent growth characteristics under stress. Furthermore, we demonstrate a prevailing trend towards negative epistatic interactions when multiple gene perturbations are combined simultaneously, thereby posing an intrinsic constraint on gene expression underlying adaptive trajectories. Together, these results emphasize how

CRISPR-Cas9 can be employed to engineer gene expression changes which shape bacterial adaptation, and present a novel approach to synthetically control the evolution of antimicrobial resistance.

5.2 Introduction

As bacteria continue to demonstrate their ability to adapt to a broad range of antibiotics[228] and other antimicrobials[229], a dearth of effective treatments for life-threatening pathogenic infections has become a prominent concern[14]. Although genomic divergences (i.e. mutations) have been the focus of conventional adaptive evolutionary research, the impact of variations in gene expression on microbial evolution during stress exposure[130,230,231] is a relatively unexplored field. Heterogeneous gene expression has been shown to enable bacterial bet-hedging[111] strategies to create diversity in order to dynamically respond to sudden environmental stressors[130]. This mutation-independent process, known as adaptive resistance[116], could expedite the evolution of antimicrobial resistance. Supporting this notion is the observance of distinct changes in bacterial transcriptomes during exposure to antibiotics[232] and disinfectants[233], as well as significant heterogeneity in inter-population gene expression during the first hundred or so generations of adapting bacterial populations[16,154,234].

In this study, we take inspiration from these adaptive strategies found in nature, and hypothesize that synthetically inducing small perturbations in gene expression can enable artificial control over both positive and negative fitness phenotypes in adapting strains. Assuming that gene expression is normally distributed around basal levels in a bacterial population, we hypothesize small changes in the distribution of gene expression

could exacerbate the pre-existing growth and fitness phenotypes of sub-populations with altered gene expression (Fig 1A). Further, we hypothesize that the simultaneous perturbation of multiple genes can induce unique phenotypic responses via epistatic interactions. Negative epistatic interactions, where the combined fitness benefits of simultaneous mutations are less than expected, have been shown to either overshadow positive epistasis during adaptation[235,236] to environmental conditions or impact long-term evolvability[237]. While it has been suggested that the epigenetic epistatic interactions of gene expression ultimately constrain long-term evolution[217], very little is understood regarding how these interactions might impact the early stages of adaptive resistance.

To investigate our hypotheses, we engineered deactivated CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-associated protein 9 (dCas9) based genomic devices to synthetically induce small perturbations in the transcriptome of *Escherichia coli* (*E. coli*). This presents a novel application of CRISPR technology as we employ it to explore the impact of subtle gene expression changes on bacterial fitness in the presence of sub lethal levels of stressors, and to the best of our knowledge is the first of its kind[238]. dCas9 and dCas9 constructs fused with the ω-subunit of RNA polymerase (dCas9-ω) have been shown to controllably inhibit[44] or activate[45] gene expression respectively. When combined *in vivo* with small guide RNAs (sgRNAs), these devices exhibit highly specific and localized control over the transcription rates of individual genes. Moreover, these CRISPR devices are able to perturb expression of multiple genes simultaneously, thereby allowing for the investigation of combinatorial effects[44] of targeted gene control and the subsequent interactions this induces.

We chose to investigate seven stress-response genes, whose functions are outlined in Supplementary Table S1. These include the global transcriptional regulators *marA*, *soxS* and *recA.* MarA (multiple antibiotic resistance) activates expression of the *mar* operon to increase efflux activity, decrease porin expression, and regulate other biochemical processes to confer tolerance to solvents and drugs[239]. SoxS (superoxide stress response) shares 49% homology in binding sites with MarA, and regulates similar genes to promote antimicrobial tolerance[240]. RecA activates the SOS response, wherein DNA repair occurs and cell growth is arrested[241]. The remaining four genes we chose to examine were downstream genes of these global regulators: *mutS*, *dinB*, *acrA* and *tolC*. MutS functions in DNA mismatch repair pathways (thereby decreasing mutation rates)[242], while DinB acts as an error-prone polymerase lacking proofreading capacity (thereby increasing mutation rates)[243]. Finally, TolC and AcrA work in tandem to construct an efflux pump to channel toxic materials outside of the cell[213]. We engineered CRISPR devices to systematically inhibit and activate the expression of these stress-response genes in *E. coli* during short-term (72 hour) exposure to various stress conditions, including antibiotics (tetracycline and rifampicin), disinfectants (bleach and hydrogen peroxide) and glucose limitation. We monitored the resulting growth and fitness impacts during the early stages of adaptation, as well as the epistatic interactions induced by simultaneous gene perturbation.

Corroborating our hypothesis, we observe that CRISPR-Cas9 based synthetic devices enable small perturbations in gene expression that are sufficient to significantly influence native bacterial adaptive responses to stress by altering growth rates, lag times, and overall fitness. We show that these impacts are reversible upon stress removal,

indicating their non-genetic nature. We demonstrate that simultaneous perturbations predominately induce negative epistasis, extending mutation-based epistasis concepts to the gene expression landscape. This work builds upon landmark gene knockout[244], plasmid over-expression[245], network rewiring[246] and long-term evolution[247] studies by outlining a novel synthetic biology approach for engineering control over bacterial adaptation via exogenously regulating gene expression profiles. Our study also helps to elucidate the early adaptive response preceding genome modifications, and serves as a paradigm shift in the field of antibiotic resistance research away from investigating downstream adaptations and towards pathways bacteria utilize for adaptation.

5.3 Results

5.3.1 Design and characterization of single target gene perturbation devices

To accomplish controlled gene expression perturbation, we designed and synthesized (see Methods) a set of 14 Type II CRISPR sgRNA plasmid constructs to inhibit or activate transcription of seven stress-response genes in *E. coli,* chosen for their known influence on adaptation[213,239–243] (Fig 1B and Supplementary Fig S1). The sgRNA constructs were named pPO-*gene*i or pPO-*gene*a for inhibition and activation respectively of each given gene (Supplementary Table S2), and were co-transformed alongside a separate plasmid containing anhydrotetracycline (aTc) inducible dCas9 or dCas9-ω into *E. coli* strain MG1655. This produced 14 unique experimental perturbation strains, designated MG1655-*gene*i or MG1655-*gene*a. Two control strains harboring dCas9 or dCas9-ω plasmids, as well as the control sgRNA construct sgRNA-RFPi (targeting the *rfp* coding sequence not present in MG1655) were also created (Supplementary Table

S3). All sgRNAs utilized common promoter and scaffolding elements, but differed in their unique 20 nucleotide (nt) sequence-specific DNA-binding domain. Inhibition and activation sgRNAs were coupled *in vivo* with dCas9 or dCas9-ω respectively to form the final protein-RNA hybrid construct with inherent DNA-binding affinity for the 20 nt sequences of each sgRNA, allowing for specific control of gene expression (Fig 1C). Activation sgRNAs targeting ≈80-110 nt upstream of the +1 transcription start site of each gene provided optimal spacing for RNA polymerase to bind to the promoter and increase gene expression[45]. Inhibition sgRNAs targeted within the first ≈50 nt of the genes' open reading frame (ORF) to inhibit transcriptional read-through via a roadblock mechanism[248]. Each CRISPR target sequence was flanked by an "NGG" Protospacer-Adjacent-Motif (PAM) on the 3'end for proper binding of the protein-RNA complex with the target DNA[248]. The impact of a subset of these constructs on neighboring genes' expression was quantified and was found to be either absent or minimal (Supplementary Fig S2). It is expected that perturbing each of these genes may induce changes in expression of downstream genes as governed by the connections through respective gene regulatory networks within *E. coli.*

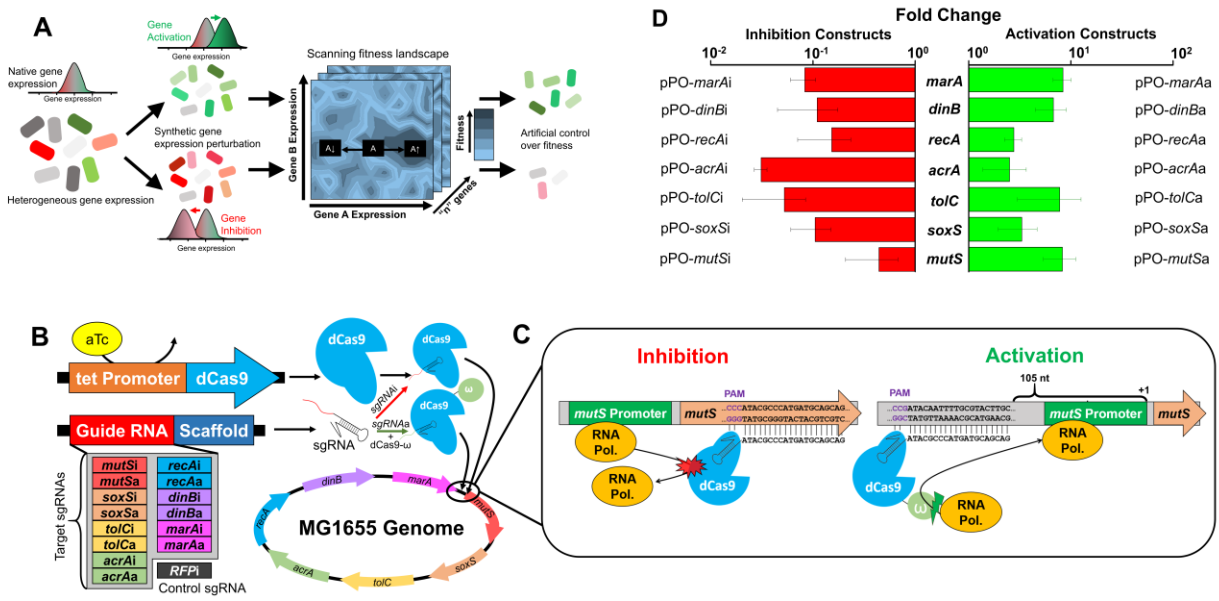


Figure 5.1 Design and characterization of synthetic CRISPR constructs perturbing gene expression

(A) Schematic demonstrating approach to engineer control over the theoretical bacterial fitness landscape[218]. By synthetically perturbing an individual gene's native expression by increasing (A↑) or decreasing (A↓) expression using synthetic CRISPR-Cas9 based genetic devices, unique fitness responses can be derived. This approach can further be applied to perturb multiple genes simultaneously to dynamically explore this adaptive landscape in n dimensions. (B) Schematic illustrating approach used to perturb gene expression in E. coli. Co-transformation of dCas9 (or dCas9-ω) plasmids with unique inhibition (or activation) sgRNA plasmids determined localization of proteins on the native E. coli MG1655 genome. These proteins combined in vivo with sgRNAs to individually perturb MG1655's expression of the seven genes shown. Note that constructs combined with the sgRNA targeting RFP did not bind to the genome, and thus served as controls. (C) Binding positions of mutS inhibition and activation constructs are shown. Inhibition constructs prevented RNAP read-through of the target's ORF, while activation constructs recruited RNAP to the promoter region by binding upstream of the +1 sequence. (D) RT-qPCR characterization of gene expression in MG1655 strains harboring dCas9 (inhibition constructs) or dCas9-ω (activation constructs) and sgRNA plasmids (pPO-genei/a), normalized to housekeeping gene rrsA and relative to wild-type MG1655. genei and genea indicate inhibition and activation respectively of the specific gene. Error bars represent standard deviation (s.d.) of biological triplicates.

The ability of bacteria to evolve resistance depends on the accessibility of higher-fitness states within a hypothetical “adaptive landscape”, which can be visualized as a multi-dimensional space comprised of the variable expression states of *n*-by-*n* genes (analogous to similar adaptive landscapes based on gene mutations)[249–251] (Fig 1A). Cloning our library of synthetic CRISPR devices into MG1655 enabled us to engineer a set of strains in which this adaptive landscape was perturbed. By inhibiting or activating individual genes, these strains enabled exploration of the impact that gene expression has on stress response. An advantage of using CRISPR devices is that this approach does not directly modify the wild-type genome, allowing for investigation of adaptive pathways in their natural state without the need to create a unique genome for each gene studied as done in canonical gene knockout studies, and thereby provides a unique insight.

To measure the effects of gene perturbation, we utilized RT-qPCR to quantify the gene expression of each of these strains relative to wild-type MG1655. Our results indicate that the strains’ expression profiles were indeed perturbed as intended, with a range of 32-fold reduction to 8-fold increase in gene expression (Fig 1D). Optimization of expression perturbation was influenced by native gene orientation; for instance, binding of dCas9-ω upstream of the +1 *soxS* transcription start site site necessitated overlap with the ORF of *soxR*, an activator of *soxS*. Growth tests were also performed to analyze the viability of these strains. No loss of viability that is not intrinsic to growth with two plasmids was observed (Supplementary Fig S3). Since MG1655-*rfp*i and MG1655-*rfp*a strains demonstrated similar growth characteristics, we used MG1655-*rfp*a as the control strain

in subsequent stress-exposure experiments (referred to hereafter as the MG1655-Control).

5.3.2 Perturbation of gene expression influences bacterial growth characteristics during stress exposure

We sought to examine the growth of strains harboring the CRISPR constructs under various environmental stresses to which infectious bacteria are commonly exposed, to determine whether artificial perturbation of gene expression enabled control over bacterial growth (and thus adaptive potential). To achieve this, five stress conditions were selected representing oxidizing agents (household bleach[252] and hydrogen peroxide[253]), antibiotics (tetracycline[254] and rifampicin[255]), and nutrient limitation (M9 minimal media supplemented with 0.4% glucose). The Minimum Inhibitory Concentration (MIC) was determined using MG1655-Control to estimate the appropriate starting concentrations for growth under each stress condition (Supplementary Fig S4). The sub-MIC levels were used as starting points for stress exposure experiments (Fig 2A**,** see Methods). We exposed *E. coli* strains harboring the CRISPR constructs to each stress over a course of 72 hours (Fig 2B), transferring biological triplicates every 24 hours into fresh media supplemented with aTc and antibiotics to maintain plasmid selection (see Methods). During this time, optical densities were monitored to track changes in growth rate (μ) and inverse lag phase (τ^{-1}) on each day of the experiment (Extended Dataset). These data was normalized to MG1655-Control by dividing μ and τ^{-1} by the average performance of biological triplicates of MG1655-Control from the experimental day (creating μ_{norm} and τ^{-1}_{norm}). Normalized data was averaged over three experimental days.

Adapting bacterial populations have been shown to exhibit significant heterogeneity in growth rates[256] and lag times[257], and thus these serve as useful metrics to quantitatively compare adaptive trends between strains. We chose to keep lag times in their reciprocal format, as larger lag times (smaller inverse lag time, τ^{-1}_{norm} <1.0) indicate that cells are stalling longer before growth and are thus considered detrimental. We found that the overall correlation between τ^{-1}_{norm} and μ_{norm} was negligible (Pearson's correlation coefficient, r = 0.09, *F*-value = 0.69), indicating that these independently provided insight into changes in growth caused by gene perturbation (Supplementary Fig S5).

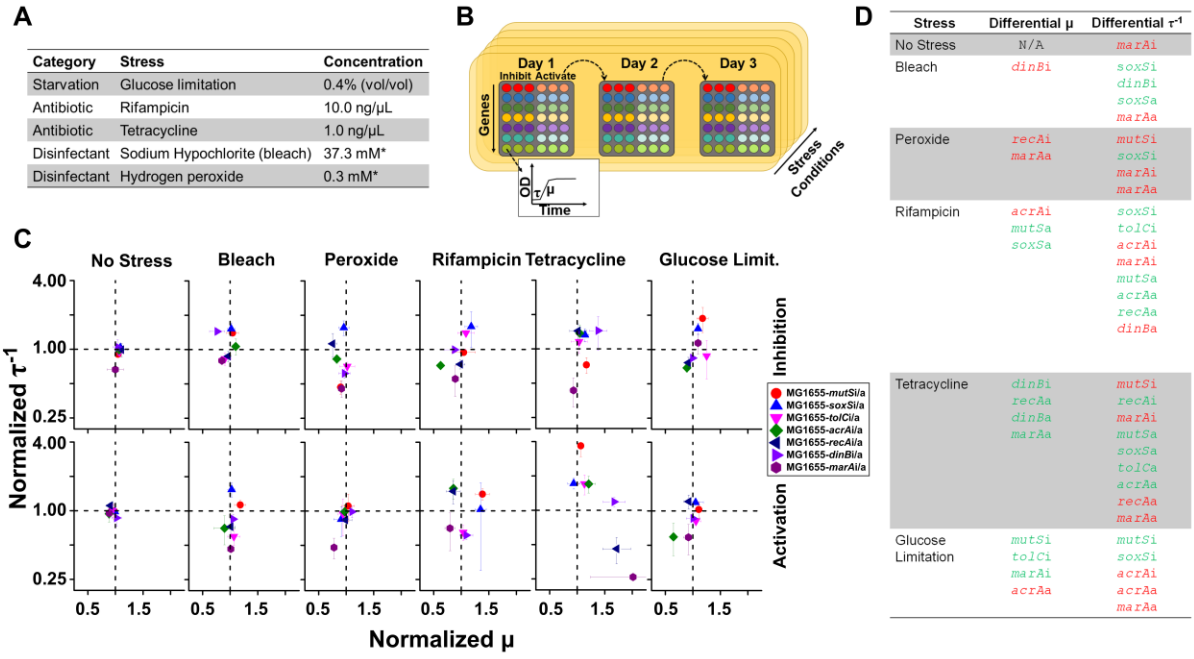


Figure 5.2 Gene perturbation during stress exposure induces altered growth characteristics

(A) Starting sub-MIC levels of toxins and nutrients for the various stress conditions. (B) Experimental design of stress-exposure while perturbing gene expression. Gene perturbation strains were exposed to individual stress conditions with serial dilution into fresh media after 24 hours exposure. OD was measured to determine τ and μ on each day. (C) Comparison of μnorm and τ-1norm averaged over three days, normalized to MG1655-Control. Deviations from intersecting dotted lines (control) indicate impacts on growth characteristics induced by perturbing gene expression with respect to MG1655-

Control. The top row shows results from inhibition strains, while the bottom shows results from activation strains. Note, the y-axis uses Log2 scaling. (D) Compiled list of results that were statistically significant at $P < 0.05$. Improvements in growth characteristics (μnorm and τ-1norm) are denoted in green; impairments are denoted in red. P-values were calculated using one-way ANOVA with Tukey post-hoc HSD test (using $\alpha < 0.05$). Error bars represent s.d. of biological triplicates.

A two-dimensional analysis of normalized τ^{-1} and μ revealed greater diversity during stress exposure than was observed under no stress (Fig 2C). From these data, 14 of the 84 growth rates (14 perturbation constructs × 6 growth conditions) and 31 of the 84 lag times demonstrated statistically significant shifts from MG1655-Control (Fig 2D). With the exception of MG1655-*marA*i, none of these shifts occurred in the absence of stress exposure, indicating that perturbations of these genes did not inherently diminish or enhance bacterial growth in absence of stress. Calculating the sum of distances (D_i and D_a for inhibition and activation constructs respectively) from the expected performance under no perturbation ($\mu_{norm} = \tau^{-1}_{norm} = 1.0$) revealed relatively minor changes under no stress ($D_i = 0.87$, $D_a = 0.63$) than under the exposure to bleach ($D_i = 2.11$, $D_a = 2.46$), peroxide ($D_i = 2.84$, $D_a = 1.24$), glucose limitation ($D_i = 2.60$, $D_a = 1.82$), and especially the antibiotics rifampicin ($D_i = 2.40$, $D_a = 3.12$) and tetracycline ($D_i = 2.81$, $D_a = 7.70$). Notably, these results indicate that shifts in growth characteristics from the control strain (deviations from the dotted lines) increased significantly during the presence of stress, demonstrating the potential that synthetically engineered gene perturbations have to artificially control the adaptive response.

Performance of perturbation strains during exposure to oxidizing agents resulted in reduced growth rates than was observed under other conditions, accounting for three of the five statistically lower μ phenotypes. Under bleach exposure, MG1655-*dinB*i

demonstrated reduced μ_{norm} (0.76 ± 0.14, P = 0.020) and increased τ^{-1}_{norm} (1.44 ± 0.05, P = 0.016), while under peroxide exposure MG1655-*marA*a demonstrated both reduced μ_{norm} (0.78 ± 0.04, P = 0.039) and τ^{-1}_{norm} (0.48 ± 0.10, P = 0.048). Lag times in particular were impacted by gene perturbations during exposure to antibiotics, affecting 8 of 14 strains grown in rifampicin and 9 of 14 strains grown in tetracycline. Under rifampicin stress, MG1655-*acrA*i demonstrated impaired growth characteristics (μ_{norm} = 0.62 ± 0.07, P = 0.038 and τ^{-1}_{norm} = 0.72 ± 0.05, P = 0.021), while MG1655-*mutS*a showed opposite effects (μ_{norm} = 1.38 ± 0.12, P = 0.044 and τ^{-1}_{norm} = 1.40 ± 0.16, P = 0.031). Under tetracycline stress, both MG1655-*recA*a and MG1655-*marA*a demonstrated improved growth rates (μ_{norm} = 1.72 ± 0.26, P = 0.022 and μ_{norm} = 2.01 ± 0.77, P = 0.0010 respectively) and extended lag times (τ^{-1}_{norm} = 0.46 ± 0.12, P = 0.022 and τ^{-1}_{norm} = 0.26 ± 0.02, P = 0.0077 respectively). Under glucose limitation, MG1655-*soxS*i showed improved growth characteristics (τ^{-1}_{norm} = 1.52 ± 0.19, P = 0.012), while MG1655-*acrA*a demonstrated the opposite effect (μ_{norm} = 0.65 ± 0.06, P = 0.0014 and τ^{-1}_{norm} = 0.59 ± 0.19, P = 0.037). Overall, these results corroborate the hypothesis that small artificial perturbations in gene expression during stress exposure significantly influence native bacterial adaptive responses.

Notably, gene perturbation under each stress condition produced unique results, indicating a complex underlying cellular response network unique to each stress. Furthermore, inhibition and activation of the same gene did not necessarily induce antagonistic effects under the same stress. For example, inhibition of *dinB* decreased μ during bleach exposure, but both inhibition and activation of *dinB* increased μ in tetracycline. Another intriguing result we observed was that both MG1655-*marA*i and

MG1655-*marA*a demonstrated significant increases in τ (under all stresses besides bleach and glucose limitation for MG1655-*marA*i and under all stresses but rifampicin for MG1655-*marA*a). This was most pronounced under tetracycline exposure (τ^{-1}_{norm} = 0.43 ± 0.12, P = 0.0029 and τ^{-1}_{norm} = 0.26 ± 0.02, P = 0.011 for inhibition and activation respectively). A correlation between over-expression of *marA* and an increase in lag time has previously been reported[258]. Conversely, MG1655-*soxS*i demonstrated reduced τ under all stresses (with the exception of tetracycline stress, where P = 0.084, all were statistically significant), while MG1655-*soxS*a also demonstrated reduced τ under bleach and tetracycline stress. Increased (or decreased) τ could indicate activation (or suppression) of persistence, which bacteria are known to employ as a survival strategy during sudden stress exposure[257]. This serves as an interesting avenue for future research into the relation between *marA* and *soxS* gene expression perturbation and the induction of persistence.

5.3.3 Competition assay confirms fitness impacts of gene perturbation

We hypothesized that if growth characteristics improved (or diminished) as a result of induced gene perturbation, then such perturbations should also provide a competitive advantage (or disadvantage) impacting bacterial fitness. To test this hypothesis, we performed competition assays between a select subset of MG1655-*gene*i/a CRISPR perturbation strains based on their phenotypic performances under stress, as well as a new control strain MG1655-mCherry. The MG1655-mCherry strain was analogous to MG1655-Control, but also included the coding sequence for mCherry on the sgRNA-RFPi plasmid. By mixing MG1655-mCherry with strains of interest, we could determine the

relative fitness impacts of gene perturbation during stress exposure utilizing flow-activated cell sorting (FACS) (Fig 3A). A mixture of the two strains grown under stress was analyzed before (D0) and after one day (D1) of stress exposure. The fluorescence of the total population was used to determine the relative ratios of the control strain with basal levels of gene expression (which fluoresced red due to the presence of mCherry) to the strain with perturbed gene expression (which did not fluoresce due to the absence of mCherry). Pure (100%) MG1655-mCherry and MG1655-Control populations distributed into two distinct fluorescence intervals both on D0 and D1 (Fig 3B). When mixed equally (50% by OD), statistically significant selection for either MG1655-mCherry or MG1655-Control was not observed after one day of exposure to tetracycline or rifampicin when compared to no stress condition (Fig 3C).

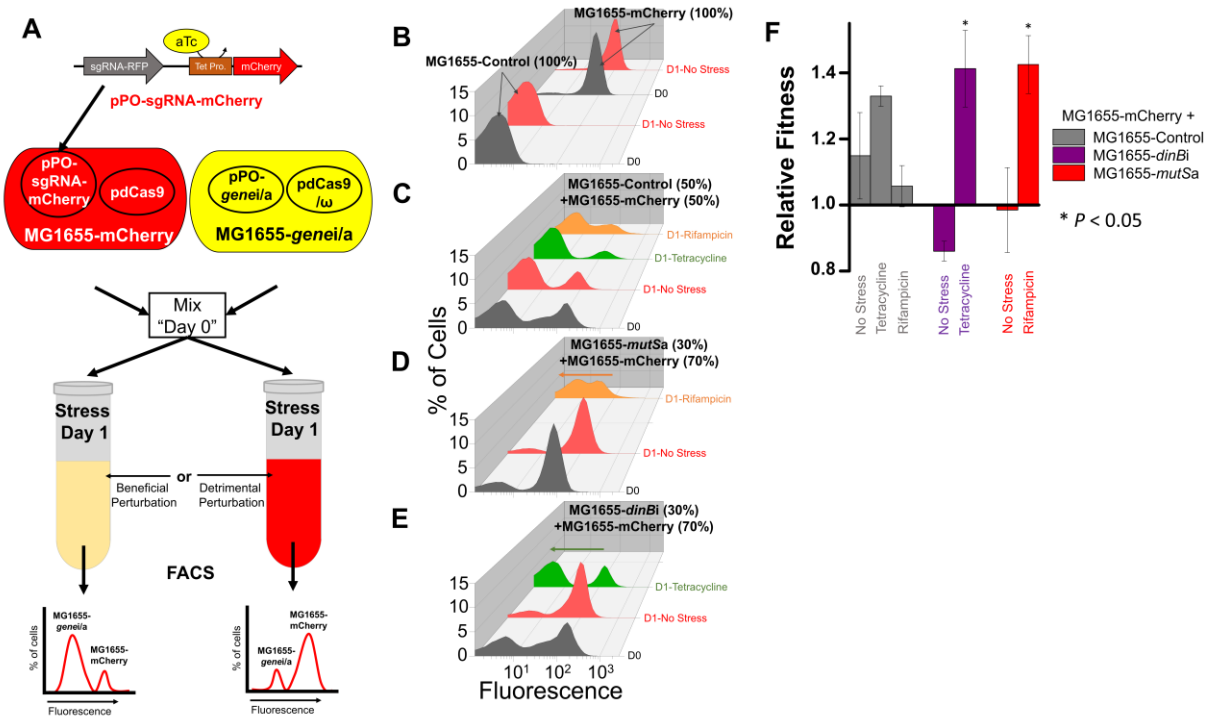


Figure 5.3 Competition assays reveal changes in fitness resulting from gene perturbations

(A) The protein mCherry, which fluoresces red, was cloned onto the sgRNA-RFPi plasmid and co-transformed into MG1655 with pdCas9 to create a fluorescent strain with basal gene expression. This strain could be distinguished from experimental MG1655 strains with perturbed gene expression via its greater fluorescence, as determined by FACS. (B-E) Representative set of FACS data from competition experiment. All day one populations (D1) were inoculated from the same day zero population (D0) in their respective graphs. (B) FACS analysis of pure MG1655-Control and MG1655-mCherry strains before (D0) and after (D1) one day of growth under no stress. (C) Control FACS analysis experiment results. MG1655-Control and MG1655-mCherry were mixed at equal starting ODs, and grown for one day under no stress (red), tetracycline (green) or rifampicin stress (orange), showing no discernable selection for either strain. (D) FACS analysis revealed selection of MG1655-*mutS*a when competed against MG1655-mCherry under rifampicin stress, despite having a lower starting concentration (30% by OD). (E) Likewise, FACS analysis revealed selection of MG1655-dinBi when competed against MG1655-mCherry under tetracycline stress, despite having a lower staring concentration (30% by OD). (F) Relative fitness levels of each strain in relation to the MG1655-mCherry strain, showing significant changes only during stress exposure across biological triplicates. *P*-values were calculated using a two-tailed type II *t*-test, and are in relation to the no stress-condition for each competition experiment. Error bars represent s.d. of biological triplicates.

To demonstrate that the MG1655-*gene*i/a strains impacted bacterial fitness during stress exposure, we utilized this competition assay approach to compare the fitness of MG1655-*mutS*a under rifampicin and MG1655-*dinB*i under tetracycline against MG1655-mCherry. These strains were selected for their measured impacts on μ shown in Figure 2. Because these strains improved μ in rifampicin or tetracycline, we chose to reduce their starting concentration (30%) relative to MG1655-mCherry (70%) on D0. As expected, activation of *mutS* and inhibition of *dinB* caused a shift in the relative population density away from MG1655-mCherry and towards MG1655-*mutS*a (Fig 3D) or MG1655-*dinB*i (Fig 3E) after one day of exposure to their respective stresses, but not during the absence of stress. The same trend demonstrated in Fig 3D was observed during competition of MG1655-*mutS*a against MG1655-mCherry using equal starting concentrations in two different growth media (LB and M9 minimal media), indicating that these results were

independent of the starting competition ratio as well as the media chosen (Supplementary Fig S6). These results were reproduced across biological triplicates, demonstrating statistically significant improvements in fitness under stress exposure (Fig 3F). We also observed the opposite effect; when MG1655-*acrA*i (70%), which exhibited reduced μ under rifampicin stress (Fig 2C-D), was competed against MG1655-mCherry (30%), the latter was selected for despite having a lower starting concentration (30%) when exposed to rifampicin stress (Supplementary Fig S7).

Furthermore, this method was utilized to estimate the fitness of each strain relative to MG1655-mCherry (Fig 3F). FACS data was used to estimate the proportion of red and non-red cells before and after one day of stress exposure, from which Malthusian (*m*) parameters were calculated for each of the two competing strains[227]. The *m* ratios were used to calculate relative fitness (see methods) from three biological replicates. The relative fitness of MG1655-Control was not statistically different between antibiotic exposure and no stress exposure conditions. However, the relative fitness of MG1655-*dinBi* was greater under tetracycline stress (1.41 ± 0.11, $P = 0.00007$), while MG1655-*mutS*a was greater under rifampicin stress (1.43 ± 0.09, $P = 0.02$) demonstrating that these strains were selected for over MG1655-mCherry only during stress exposure. Overall, these results corroborate the measured fitness impacts on growth characteristics as shown in Fig 2C-D.

5.3.4 Phenotypic reversibility and evidence of non-genetic impact of transcriptional perturbation

We next tested the reversibility of the phenotypic deviations of these experimental strains from MG1655-Control; that is, whether or not the CRISPR-perturbed strains demonstrated similar growth characteristics as the control upon removal of stress. Such reversion under no stress would highlight the non-genetic nature of the observed changes in growth characteristics and fitness. We performed such analysis on a subset of biological triplicates collected at the end of three days of exposure to tetracycline and rifampicin stress (Fig 4A). These strains were grown for one day in the absence of stress and aTc induction. Afterwards, each strain was re-exposed to the three day adaptation experiment in the absence of stress both with and without aTc induction, or the same initial stress and aTc. For rifampicin adapted strains, we observed a return to the wild-type phenotype in no stress both in the absence (D_i = 0.60, D_a = 0.96) and presence (D_i = 0.99, D_a = 0.75) of aTc induction of gene perturbation (Fig 4B). A similar phenomenon was observed in tetracycline adapted strains under no stress in the absence (D_i = 0.90, D_a = 1.08) and presence (D_i = 1.24, D_a = 0.66) of aTc (Fig 4C). Together, these data indicate that the phenotypic effects of gene expression perturbations were reversible upon stress removal, suggesting their non-genetic nature.

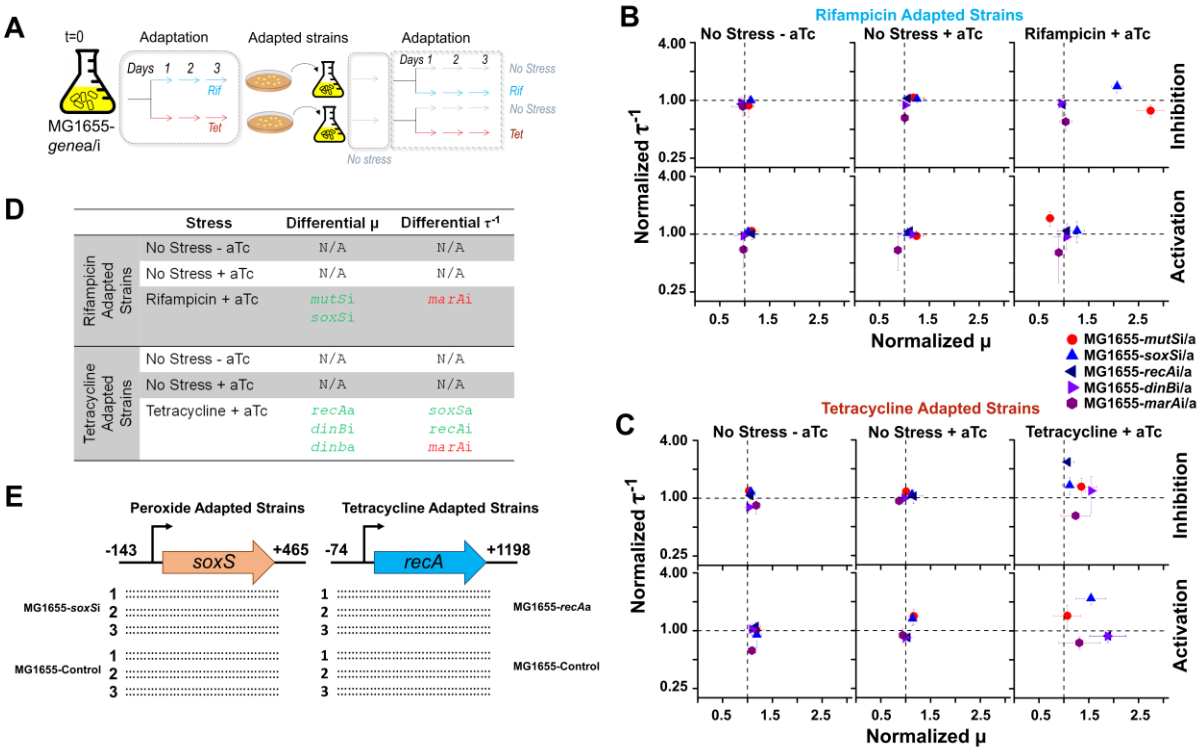


Figure 5.4 Evidence of phenotypic reversibility of adapted strains

(A) Schematic illustrating approach to investigate reversibility of phenotypic changes observed during growth under stress. Strains collected after 72 hours of stress exposure were plated, grown for 24 hours under no stress, and exposed to either no stress (with and without aTc induction) or the initial stress they were exposed to. (B) Strains exposed to rifampicin stress were collected and grown for another three days under either no stress or rifampicin exposure. Perturbation strains behaved similarly to MG1655-Control once returned to no stress, but maintained their unique phenotypes under rifampicin exposure. Error bars represent s.d. of biological triplicates. (C) A similar phenomenon was observed by strains which were exposed to tetracycline stress, collected, and grown for another three days under either no stress or tetracycline exposure. (D) Statically significant deviations in μ and τ^{-1} of rifampicin and tetracycline adapted strains after re-exposure to stress, using a one-way ANOVA with Tukey post-hoc HSD test (using $\alpha < 0.05$). (E) After exposure to peroxide (MG1655-soxSi and MG1655-Control) and tetracycline (MG1655-recAa and MG1655-Control), biological duplicates of each biological triplicate were sequenced around the genetic region where synthetic perturbation constructs targeted. Sequenced regions are indicated with respect to the +1 of the promoter. No mutations were observed in any of the 24 strains. Dots indicate change with respect to the wild-type.

When stress was maintained, perturbation strains continued to demonstrate deviations in μ and τ^{-1} under rifampicin exposure (D_i = 3.49, D_a = 1.38) as well as

tetracycline exposure (D_i = 3.18, D_a = 3.93). A majority of strains exhibited similar phenotypes during both the first and second round of rifampicin exposure. MG1655-*marA*i again demonstrated a reduced τ^{-1}_{norm} (0.66 ± 0.08, P = 0.038). The previously identified lag time impacts of MG1655-*recA*i, MG1655-*recA*a and MG1655-*dinB*a became less pronounced. Two new phenotypes were observed only during the second round of rifampicin exposure: MG1655-*mutS*i μ_{norm} (2.74 ± 0.27, $P = 6.9 * 10^{-8}$) and MG1655-*soxSi* μ_{norm} (2.07 ± 0.11, $P = 1.2 * 10^{-5}$). The difference in phenotypes between the first and second rounds of adaptation could be explained by an altered epigenetic state over an extended period of adaptation.

Under the second round of tetracycline exposure, no such new phenotypes were observed. Six previous statistically significant results remained; MG1655-*recA*a, MG1655-*dinB*i and MG1655-*dinB*a exhibited increased μ_{norm} (1.89 ± 0.35, P = 0.041, 1.55 ± 0.11, P = 0.0040 and 1.88 ± 0.36, P = 0.042), MG1655-*soxS*a and MG1655-*recA*i exhibited increased τ^{-1}_{norm} (2.18 ± 0.35, P = 0.0084 and 2.35 ± 0.08, P = 0.0019 respectively), and MG1655-*marA*i exhibited decreased τ^{-1}_{norm} (0.65 ± 0.03, P = 0.042).

We next sequenced the experimental strains for mutations. Both MG1655-Control and gene perturbation strains received the same basal level of selection pressure to accumulate alterations at the genomic level. However, it is theoretically possible that CRISPR perturbations could have artificially induced mutations in their genomic targets to circumvent the synthetically induced gene expression changes, undermining the observed phenotypic changes in perturbation strains. To test for this possibility, we chose to sequence the genes directly influenced by perturbation in a subset of our strains (Fig 4E). We sequenced the region of *soxS* in MG1655-*soxS*i after exposure to peroxide

stress, as well as the region of *recA* in MG1655-*recA*a after exposure to tetracycline stress, since these gene perturbations showed significant impact on μ and τ^{-1} respectively. We also sequenced MG1655-Control after exposure to each condition, to account for any mutations not influenced by gene perturbation. Sequencing of six biological replicates per strain provided no evidence of mutations, indicating that these perturbations likely did not induce mutations in the genomic regions they targeted. However, these results do not discount the possibility of other mutations arising in experimental strains and not in MG1655-Control or vice-versa. It is expected that the gene perturbations could have influenced other genomic regions to mutate during stress exposure.

In order to characterize the consequence of gene perturbations on mutation rates across the genome, we performed a mutation fluctuation assay on a subset of strains: MG1655-Control, MG1655-*mutS*i, MG1655-*mutS*a, MG1655-*soxS*i, MG1655-*soxS*a, MG1655-*dinB*i, MG1655-*dinB*a and MG1655-*mutS*i-*dinB*a (Supplementary Fig S8). Our results indicated no significant difference in mutation rates between MG1655-Control and all strains tested, with the noticeable exception of MG1655-*soxS*a which was shown to have a significantly increased mutation rate over MG1655-Control in the absence of stress (1.6×10^{-8} versus 5.1×10^{-10} respectively). One possibility for the increased mutation rates of MG1655-*soxS*a is an increased expression of Endonuclease IV gene *nfo*, one of the genes regulated by SoxS[259]. These results indicate that while CRISPR perturbation constructs did not cause mutations in their gene targets, they did potentially alter mutation rates across the whole genome and could serve to explain some of the altered phenotypes during stress exposure. Combined with the observed phenotypic-reversal of

gene expression perturbations upon stress removal and the lack of mutations in genomic regions directly targeted by CRISPR constructs, these results point towards the fact that the observed changes in the fitness phenotypes are mainly an effect (via altered mutation rates or otherwise) induced by the CRISPR devices, and are unlikely due to independent genetic changes that arose in absence of CRISPR perturbation influence.

5.3.5 Design and characterization of strains perturbing multiple targets simultaneously

We next explored the combinatorial effects of up-regulating or down-regulating multiple genes simultaneously with CRISPR constructs. In principle, combining several independent sgRNA targets into one construct allows for controlled perturbation of multiple genes' expression patterns, as has recently been demonstrated[225]. This provides a powerful tool to tune expression of multiple genes (either increase or decrease expression) without the need to use multiple artificial promoters that require to be integrated in the genome. We demonstrated simultaneous induction of synthetic gene perturbations using a modified cloning approach that introduced tandem independent sgRNAs onto one plasmid to combine *in vivo* with dCas9 or dCas9-ω (Supplementary Fig S9). We designed three constructs in which the perturbed genes had known regulatory interactions: MG1655-*tolC*i-*acrA*i (which inhibited the AcrAB-TolC multidrug efflux pump), MG1655-*mutS*a-*dinB*i (which activated expression of a mismatch-repair protein and decreased expression of an error-prone polymerase, thereby likely decreasing mutation rates), and MG1655-*recA*a-*dinB*a (which increased expression of *dinB* both directly and indirectly by increasing expression of its upstream up-regulator *recA*). We designed a

fourth construct, MG1655-*tolC*i-*acrA*i-*soxS*i, which inhibited three genes simultaneously which showed similar impacts on growth characteristics in Figure 2C-D (*tolC*i and *soxS*i increased μ under glucose limitation and τ under rifampicin exposure, while *acrA*i and *soxS*i increased τ under tetracycline exposure). Finally, we designed a fifth construct, MG1655-*tolC*i-*dinB*i, which perturbed two genes in separately regulated pathways and have not been shown to produce similar impacts on growth characteristics under the same stress condition.

Strains engineered to only exhibit inhibited gene expression utilized dCas9, while strains engineered to exhibit activation of one or more genes utilized dCas9-ω. Notably, strain MG1655-*mutS*a-*dinB*i demonstrates that simultaneous activation and inhibition of gene expression is possible through the use of dCas9-ω (Fig 5A). Previous work using a dCas9-VP64 eukaryotic transcriptional activator validated that CRISPR activation constructs can reliably inhibit gene expression when targeted to areas downstream of the promoter sequence[260]. We utilized RT-qPCR to verify that MG1655-*mutS*a-*dinB*i (as well as MG1655-*tolC*i-*acrA*i-*soxS*i*)* perturbed gene expression as intended (Fig 5B). Growth for 72-hours under stress was repeated for strains harboring these multi-target synthetic constructs.

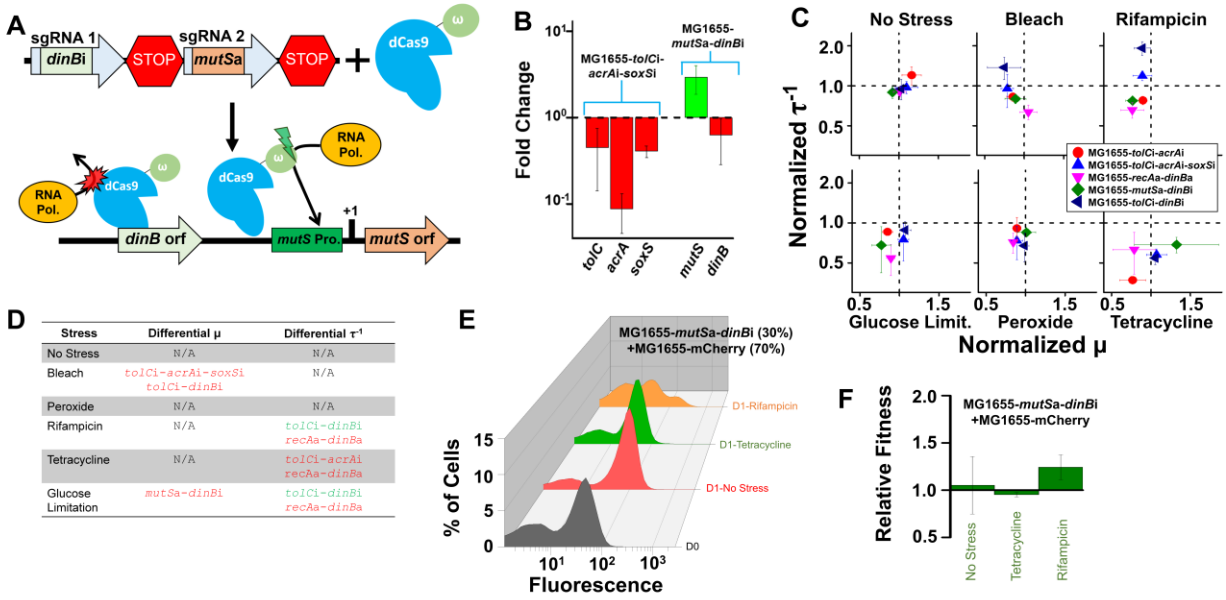


Figure 5.5 Utilization of CRISPR constructs to simultaneously perturb expression of multiple genes

(A) Strategy for perturbing multiple genes simultaneously involved cloning tandem targets into sgRNA constructs, which were transcribed individually to guide dCas9 or dCas9-ω (as demonstrated by constructs harbored by MG1655-*mutS*a-*dinB*i). (B) RT-qPCR characterization of gene expression in strains MG1655-*tolC*i-*acrA*i-*soxS*i and MG1655-*mutS*a-*dinB*i relative to wild-type MG1655, normalized to the housekeeping gene *rrsA*, revealing constructs simultaneously perturbed multiple genes' expression states as intended. (C) Impact of constructs on growth characteristics under various stress conditions, relative to MG1655-Control (dotted lines). (D) Compiled list of growth characteristics that were statistically significant at $P < 0.05$. Improvements in growth characteristics (normalized values greater than 1.0) are denoted in green; impairments are denoted in red. (E) Representative set of FACS data from competition of MG1655-*mutS*a-*dinB*i against MG1655-mCherry. Cells were mixed such that MG1655-*mutS*a-*dinB*i constituted 30% of the total starting OD. Unlike MG1655-*mutS*a or MG1655-*dinB*i, this strain was not selected for under tetracycline or rifampicin stress. (F) Fitness of MG1655-*mutS*a-*dinB*i relative to MG1655-mCherry, in which fitness under stress was not shown to be statistically different than fitness under no stress across biological triplicates. *P*-values were calculated using one-way ANOVA with Tukey post-hoc HSD test (using $\alpha < 0.05$). Error bars in (c, f) represent s.d. of biological triplicates.

The effects of modulating gene expression on growth characteristics were quantified, again normalizing data against the MG1655-Control strain grown alongside the multi-target strains (Fig 5C and Extended Dataset). We found 3 of 30 measured μ (5

perturbation constructs × 6 growth conditions) and 6 of 30 measured τ^{-1} were significantly impacted by simultaneous gene perturbations. Surprisingly, 7 of the 9 statistically different growth characteristics measured by these strains were detrimental, i.e. decreased μ_{norm} or τ^{-1}_{norm}. The only two improvements were exhibited by a higher τ^{-1}_{norm} of MG1655-*tolC*i-*dinB*i under rifampicin stress and glucose limitation (Fig 5D). Conversely, MG1655-*recA*a-*dinB*a demonstrated reduced τ^{-1}_{norm} under three different conditions: rifampicin stress (τ^{-1}_{norm} = 0.66 ± 0.09, *P* = 0.0064), tetracycline stress (τ^{-1}_{norm} = 0.63 ± 0.22, *P* = 0.0030) and glucose limitation (τ^{-1}_{norm} = 0.54 ± 0.14, *P* = 0.0081). Interestingly, the perturbation of *recA*a and *dinB*a alone had reciprocal impacts under rifampicin and tetracycline stress, and no significant impact under glucose limitation. These results indicated that the growth of strains with simultaneously perturbed genes was diminished in relation to single target strains.

As before with single-gene perturbation, we employed competition assays between MG1655-*mutS*a-*dinB*i and MG1655-mCherrry to further analyze the fitness impacts induced by simultaneous gene perturbation. We saw no selection for MG1655-*mutS*a-*dinB*i after one day of stress exposure in either rifampicin or tetracycline (Fig 5E). Competition was also performed between MG1655-mCherry and MG1655-*tolC*i-*dinB*i, which did show a shift towards the later population after rifampicin exposure as predicted by growth characteristic data (Supplementary Fig S10). Using Malthusian (*m*) parameters calculated from biological triplicates, there was no statistically significant difference between the relative fitness of MG1655-*mutS*a-*dinB*i under no stress and the relative fitness under either stress (Fig 5F). These data contrasts the improved fitness of MG1655-*mutS*a and MG1655-*dinB*i in rifampicin and tetracycline respectively that was

previously observed (Fig 3D-F). This suggests that strains in which multiple genes are perturbed are less fit than would be expected based on results from single-gene perturbation strains.

5.3.6 Simultaneous gene perturbation predominantly yields negative epistatic interactions

We next examined the epistasis induced by simultaneous gene perturbation (recently characterized as 'epigenetic' epistatic interactions[217]), given that a large number of strains harboring multiple gene targeting CRISPR constructs elicited a less-fit phenotype than predicted from the performance of strains harboring their single-target constituents. To do so, we utilized a multiplicative model to calculate epistasis in both μ_{norm} and τ^{-1}_{norm} for each of the five strains with simultaneous gene perturbations. The expected growth rates (or inverse lag times) of these strains were calculated by multiplying together the μ_{norm} (or τ^{-1}_{norm}) of each single gene perturbation strain from which they were created, and epistasis was calculated as the difference between these expected values and those that were measured (see methods).

Epistasis in μ_{norm} and τ^{-1}_{norm} was quantified for each strain under each growth condition, and the distribution of epistasis was analyzed (Fig 6A). Indeed, we observed a distinct pervasiveness of negative epistasis resulting from simultaneous gene perturbation in both μ_{norm} (mean epistasis = -0.17 ± 0.14 [95% confidence interval], P = 0.02) and in τ^{-1}_{norm} (mean epistasis = -0.35 ± 0.33 [95% confidence interval], P = 0.04). A large majority of the data (73% of μ_{norm} epistasis and 63% of τ^{-1}_{norm} epistasis) falls into the category of negative epistasis, and both distributions are markedly skewed towards

greater magnitudes of negative epistasis. These results help to explain the prevalence of diminished growth characteristics and fitness observed. The data also suggests that when two or more genes are perturbed from their basal expression levels, their combinatorial fitness benefits are generally abated or their disadvantages amplified.

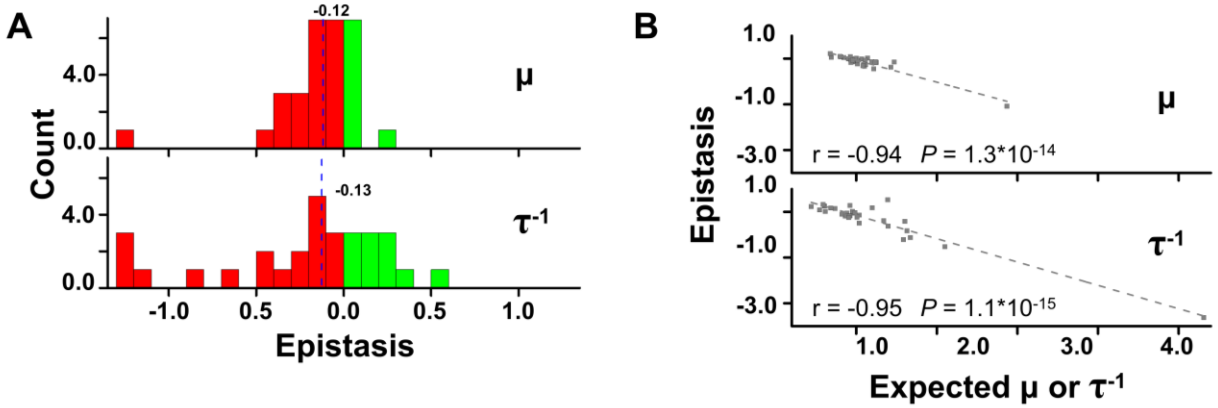


Figure 5.6 Induction of negative epistasis due to simultaneous gene perturbation

(A) Distribution of epistatic interactions on μ_{norm} and τ^{-1}_{norm} of five strains in which genes were simultaneously perturbed, over all six growth conditions. The distribution is skewed to the left, revealing the prevalence of negative epistasis resulting from simultaneous gene perturbation. Vertical blue lines indicate median values of epistasis. (B) A strong negative correlation was observed between the expected growth characteristic of the five strains and their measured epistatic interactions. Linear fits utilize epistasis data from all five strains under all six growth conditions.

An analysis of epistasis in μ_{norm} (Supplementary Fig S11) and τ^{-1}_{norm} (Supplementary Fig S12) of each strain revealed that the magnitude of epistasis depended more heavily on the stress exposure, rather than the strain itself. Clustering analysis revealed that epistatic trends appeared to group by stress (Supplementary Fig S13). This was apparent especially under tetracycline exposure, which resulted in much larger degrees of negative epistasis in both μ_{norm} and τ^{-1}_{norm}. These findings could indicate disruption of stress-dependent adaptive pathways and poses an avenue for future research.

Interestingly, negative correlations between expected growth characteristics and their measured epistasis were also observed (Fig 6B). Removing the most negative value of epistasis still resulted in significant negative correlations ($r = -0.70$, $P = 2.4*10^{-5}$ and $r = -0.79$, $P = 1.1*10^{-7}$ for μ_{norm} and τ^{-1}_{norm} respectively). This implies that negative epistatic effects are strongest whenever multiple gene perturbations, which individually prove beneficial, are subsequently combined. This trend resembles diminishing returns epistasis, wherein the fitness gains of consecutive mutations decelerate during adaptation[235,236]. This phenomenon has been reproduced across a number of studies, and has been correlated to mutations which specifically impact gene expression[218]. It has been suggested that epigenetic epistatic interactions can constrain the evolution of gene expression[217]. The data presented here suggests an inherent fitness cost is associated with excessive perturbations of gene expression levels, and that epigenetic interactions may be subjected to the same diminishing returns epistasis typically associated with mutations.

5.4 Discussion

Here we apply CRISPR technology, which thus far has been used primarily in a genome-editing capacity, to explore the effects of gene perturbations on bacterial adaptation. We demonstrate that minor (relative to previous studies[244,245]) perturbations in gene expression patterns are sufficient to reversibly alter bacterial fitness and growth characteristics during the early stages of stress exposure. We also identify significant negative epistasis when multiple genes are perturbed simultaneously. These results highlight the importance of gene expression in regulating bacterial adaptive responses

and influencing the evolutionary pathways available during the early stages of stress exposure. This study also demonstrates a novel approach to rationally engineer control over bacterial fitness (and thus adaptive potential) without directly modifying the organism of interest.

Observed changes in gene expression during stress exposure have been extensively reported in bacteria[261,262]. We have recently demonstrated that *E. coli* grown under the same stress exhibits significant inter-population heterogeneity in gene expression, where some genes demonstrate high expression variability while others are constrained within specific signatures[16]. This complexity surrounding differential gene expression during evolution demonstrates that bacteria have many avenues to explore their adaptive landscapes and develop novel, mutation-independent strategies to survive under stress[126]. Corroborating this notion is our finding that the phenotypic changes induced by gene perturbation during stress exposure are reversible upon the removal of stress, suggesting the importance of gene expression to the adaptive processes. It is important to note that all 19 strains of *E. coli* started from the same isogenic background (although harbored different CRISPR constructs) before exposure to stress. After exposure to stress, gene perturbations likely influenced the acquisition of mutations (as suggested by the increased mutation rates of MG1655-*soxSa* over MG1655-Control even in the absence of stress), resulting in heterogeneous end-populations. These mutations, however, were either not costly under normal conditions, or a compensatory mutation arose to accommodate their deleterious effects. We do not deny these possibilities; rather, we argue that any mutations that arose (or failed to arise) in one population but

not another could have been influenced by the unique synthetic gene perturbation we applied to said population.

Related efforts to study antimicrobial resistance has focused primarily on the effects of direct genetic modifications such as gene deletions or increased copy numbers via exogenous expression on a plasmid. For instance, the landmark genome-wide screen of Keio knockouts under 324 stress conditions by *Nichols et al.* enabled characterization of orphan genes' functions and connections within previously identified biological pathways[244]. Conversely, *Soo et al.* demonstrated that increasing gene copy numbers 300-400 fold utilizing the ASKA library increased the MIC in 115 cases[245]. While studies such as these and others[263] have been invaluable in relating gene function to survival under particular stresses, they employ genetic extremes (i.e. complete gene knockouts or 300-fold gene copy increase) relative to the smaller changes in gene expression induced in this study. As such, their results less accurately represent what might be found in a natural distribution of a population of bacteria. Our approach maintains the integrity of the original genome, while only introducing epigenetic changes via plasmids. The degree of gene expression perturbation is also tunable; by changing the targeting site of the sgRNA, or the amount of aTc added to the system, one can adjust the efficiency of gene repression and activation[45]. Furthermore, a number of our results are consistent with previous studies. For instance, we observed that MG1655-*soxS*i exhibited decreased lag time under all five stress conditions, which is consistent with previous observations of *soxS* overexpression during lag phase[264] and could indicate that the superoxide stress response contributes to tolerance-by-lag[257].

Interestingly, we also observed occasional phenotypic disconnect between activation and inhibition of gene targets, in that they did not impart contrasting phenotypes on adapting populations. For instance, MG1655-*dinB*i/a strains exhibited both beneficial and detrimental growth and fitness impacts, suggesting that DinB's error-prone activity is only conditionally beneficial. This is corroborated by confounding conclusions from previous studies, such as findings that *dinB*$^{-}$ exhibits increased fitness under certain antibiotics[244] but decreased fitness under others[244]. Indeed, we and others observe that perturbations (both genetic and epigenetic) impact fitness differently across a set of stresses. For instance, *acrA* and *tolC* have been shown to be up-regulated during the global stress response[265] but do not necessarily contribute to tolerance or can even promote increased fitness when knocked out[244]. We observe similar results in MG1655-*acrA*i/a, where inhibition reduced fitness in rifampicin but activation also reduced growth in nutrient starvation. A possible explanation for the disconnect between gene activation and inhibition across stresses could be that native gene expression is inherently large or small enough that skewing it further upwards or downwards respectively does not provide an adaptive advantage. This phenomenon has been postulated to explain the fact that the RpoS σ-factor does not induce *dinB* expression in response to ampicillin, as basal levels of DinB appear to be sufficient for ampicillin-induced mutagenesis[137]. However, this would not explain why both activation and inhibition of particular genes produced drastically different phenotypes over MG1655-Control. For instance, while previous studies have found *marA*$^{-}$ to decrease colony size[244] and *marA* hyper-expression to outcompete the wild-type[245] during tetracycline exposure, we find that in similar conditions both ~10 fold *marA* inhibition and activation significantly increased τ. Thus, an alternate

explanation is that gene targeting by these constructs disrupts regulatory pathways responsible for altering expression in response to environmental stress, such as the feedback network which has been postulated to control pulsation of *marA* expression[114]. This regulatory complexity, compounded by the connection of many of these genes within regulatory networks, obfuscates potential modular approaches to developing next-generation antimicrobial targets and highlights the need for studies such as this to better understand the consequences of artificial gene perturbation.

The canonical approach to study individual gene knockouts or duplications also neglects the combinatorial effects of multiple interacting genes. These effects are crucial in determining how adaptive trajectories unfold – the fixation of a mutation during the early stages of adaptation can constrain bacteria to a particular adaptive pathway that precludes another competing avenue of adaptation[266]. Previous studies have demonstrated that there is a pervasive tendency for coinciding bacterial mutations to display negative epistatic interactions[235,236]. Wright's shifting balance model predicts that epistasis plays a fundamental role in determining the likelihood of evolutionary drift overcoming fitness barriers to adaptation[249,267]. Utilizing our CRISPR-Cas9 based gene perturbation approach, we expand on this idea of negative epistasis controlling evolutionary pathways by discovering a striking correlation between concurrent perturbation of multiple genes' expression levels and negative epigenetic epistatic interactions. It should be noted that qPCR results demonstrate that there may have been a diminished gene expression impact from simultaneous gene perturbation than from individual gene perturbation, which could be a result of dCas9 (or dCas9-ω) protein being guided by multiple sgRNAs and thus diluting its concentration at individual genomic loci.

However, this would not explain why the impacts of simultaneous gene perturbation were predominantly detrimental in nature, and might even suggest that greater negative epistatic trends might be encountered if dCas9 (or dCas9-ω) expression were to be increased to counteract potential sgRNA dilution effects. Regardless, this is the first reported observation of synthetically induced negative epistasis to the best of our knowledge. Just as negative epistatic interactions derived from mutations play a critical role in shaping bacterial adaptive trajectories, negative epigenetic epistatic interactions based on variations in gene expression also appear to be fundamentally important in shaping the availability of adaptive pathways[217] and should be taken into consideration when investigating how novel resistances evolve.

5.5 Methods

5.5.1 Bacterial strains, media and culture conditions.

E. coli cloning strains NEB 10-β (New England Biolabs) and DH5α (Zymo Research Corporation), as well as the final experimental strain K-12 MG1655 (ATCC 700926) were cultured in Luria-Bertani Broth (LB) (Sigma-Aldrich®) unless otherwise noted. Colonies were grown on LB-agar plates supplemented with ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL). For nutrient limiting conditions and growth of RT-qPCR biological triplicates, M9 minimal media (5X M9 minimal media salts solution from MP Biomedicals, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$ in sterile water) was used in lieu of LB, supplemented with 0.4 % vol/vol glucose (34.2 mM). Expression of dCas9 and dCas9-ω was induced from plasmids pdCas9 and pPO-dCas9ω respectively by adding aTc (10 ng/mL) to the media. During competition experiments, aTc concentration was

increased to 25 ng/mL to assist in distinguishing fluorescent populations from non-fluorescent ones. Cloning strains were made chemically competent with the *Mix & Go E. coli* Transformation Kit & Buffer Set (Zymo Research Corporation), and the final sgRNA plasmids were transformed into electrocompotent MG1655 cells harboring either pdCas9 or pPO-dCas9ω for inhibition or activation of gene expression respectively. The final experimental strains are listed in Supplementary Table S3. Cultures (2-5 mL) were grown at 37°C with constant shaking at 225 r.p.m. All experiments used biological triplicates inoculated from individual colonies grown on LB-agar plates supplemented with ampicillin and chloramphenicol.

5.5.2 Plasmid assembly.

A list of plasmids used in this study is provided in Supplementary Table S2**.** dCas9-ω from pWJ66 (Addgene plasmid 46570)[248] was first cloned into the same vector as dCas9 from pdCas9 (Addgene plasmid 44249)[45] under the same aTc inducible promoter, rrnB T1 terminator and chloramphenicol resistance marker to create plasmid pPO-dCas9ω. Single-target sgRNA plasmids were first derived from the RFP-targeting control sgRNA-RFPi (Addgene plasmid 44251)[248] using the approach outlined in Supplementary Fig S9A. Primers were designed to replace the 44251 plasmid's RFP-targeting sgRNA using a common reverse primer (Reverse *sgRNA*) flanked with an *ApaI* restriction site and unique forward primers flanked with a *SpeI* restriction site, listed in Supplementary Table S4. PCR with Phusion® High-Fidelity DNA Polymerase (New England Biolabs) was used to amplify these new target sgRNA-insert DNA fragments, which were subsequently gel-purified (Zymoclean™ Gel DNA Recovery Kit – Zymo Research Corporation),

digested with *ApaI* and *SpeI* (FastDigest Enzymes – Thermo Scientific) as per the provided protocols and PCR-purified (GeneJET PCR Purification Kit – Thermo Scientific). The 44251 plasmid (Addgene) backbone was also digested with *ApaI* and *SpeI* and gel purified, and T4 DNA Ligase (Thermo Scientific) was used to ligate the new sgRNA target inserts into the 44251 backbone. Ligations were transformed into chemically competent DH5α or NEB 10-β cells. Plasmids minipreps were performed using Zyppy™ Plasmid Miniprep Kit (Zymo Research Corporation). Sequencing of final sgRNA constructs were performed for validation of correct assembly product (GENEWIZ).

To create the fluorescent control for competition assays, mCherry from pHL662 (Addgene plasmid 37636) was PCR amplified with *AatII* restriction sites on either end of the resulting fragment, and cloned onto sgRNA-RFPi under an aTc inducible promoter. This construct was transformed into MG1655 harboring pdCas9 to create MG1655-mCherry.

To create multi-gene targeting sgRNA plasmids, the above single gene targeting sgRNAs were combined using the procedure outlined in Supplementary Fig S9B. The first sgRNA target plasmid was digested with *BamHI* and *XbaI* and the 2569 bp vector was gel-purified. A second target plasmid was digested using *BglII* and *XbaI* and the 527 bp insert was gel-purified. These were ligated and transformed into DH5α chemically competent cells and plated on LB-ampicillin plates. *BamHI* and *BglII* digestion overhangs produce compatible sticky ends that, when ligated together, do not produce a new restriction enzyme site. After recovering these plasmids, a *BamHI* digestion was used to confirm the plasmids were the correct size on an agarose gel. To create more than two targets, the same approach was applied using the *BamHI*/*XbaI* digestion on the two gene

targeting plasmid and the *BglII*/*XbaI* digestion on the third target. All inserted fragments were confirmed by sequencing.

5.5.3 RT-qPCR analysis.

Biological triplicate cultures inoculated from individual colonies were grown with appropriate antibiotics overnight in 2.5 mL M9 minimal media. Cultures were induced with aTc the following morning for three hours, and cell pellets were subsequently collected and RNA extracted using the GeneJET RNA Purification Kit (Thermo Scientific) supplemented with lysozyme and proteinase K. Collected RNA was then purified using the TURBO DNA-free kit (Ambion) for heavy DNA contamination. cDNA was synthesized from these RNA samples using 10 µL reactions of the DyNAmo SYBR Green 2-Step qRT-PCR kit (Thermo Scientific). A control reverse-transcriptase-free reaction was included in tandem with all cDNA synthesis reactions. Technical duplicates of RT-qPCR reactions were performed in 10 µL reactions using 2 ng of cDNA and 0.5 µM primers listed in Supplementary Table S5. Primer efficiency and specificity were confirmed in a previous experiment[16]. Samples were run on an Eco Real-Time PCR System (Illumina) in the CU Core Sequencing Facility operating the Eco Software v4.1.2.0. RT-qPCR reactions of neighboring genes' expression were performed in 20 µL using 2 ng of cDNA and 0.5 µM primers, and were run on a QuantStudio™ 6 Flex Real-Time PCR System (Thermo Scientific) in the CU Core Sequencing Facility. An initial 10 min polymerase activation at 95°C was performed, followed by 40 cycles of 95°C 15 second denaturation, 55°C 30 second annealing, and 72°C 30 second extension. Rox normalization was used to compare qPCR samples, and the average C_q values of technical duplicates were used to

calculate the $\Delta\Delta C_q$ values using *rrsA* gene expression as a housekeeping gene, which was also averaged. Fold change was calculated using $2^{-\Delta\Delta Cq}2^{-\Delta\Delta Cq}$ values for individual biological triplicates, which were subsequently used to obtain averages and standard deviations.

5.5.4 MIC determination.

MICs were first determined via overnight growths of MG1655 harboring dCas9 and RFP-targeting sgRNA with no induction, and measuring the change in OD at 562 nm. A range of concentrations for the disinfectants (hydrogen peroxide and bleach) and the antibiotics (rifampicin and tetracycline) were tested, reducing the concentration by half between. MIC tests started at concentrations of 74.6 mM, 20.4 mM, 10 ng/μL and 100 ng/μL for bleach, hydrogen peroxide, rifampicin and tetracycline respectively. MICs were determined to be the lowest concentration which prevented a change of 0.1 OD between days. The final sub-MIC values used in this experiment, as well as a description of mechanisms of action, are presented Fig 2A.

Stress conditions were selected to monitor a broad range of antimicrobials. Peroxide and bleach introduce oxidative stress by producing highly reactive superoxide[253] or chlorine[252] radicals respectively. Tetracycline inhibits protein synthesis by disrupting tRNA interactions with the ribosome[254], while rifampicin inhibits transcription by preventing the activity of RNA polymerase[255]. These antibiotics avoided mechanistic overlap with the antibiotics required to maintain plasmid selection, ampicillin and chloramphenicol, which inhibit cell-wall-synthesis and peptide bond formation respectively.

5.5.5 Stress growth conditions.

Biological triplicates were inoculated from individual colonies of MG1655 cultures harboring both dCas9/dCas9-ω and sgRNA plasmids into 100 µL LB cultures supplemented with ampicillin, tetracycline and aTc and grown overnight to stationary phase. The next day, 2 µL was used to inoculate one 50 µL M9 culture and five 50 µL LB cultures in a 384-well microplate, all of which were supplemented with aTc and the appropriate antibiotics. Four of the LB cultures were supplemented with either 37.3 mM sodium hypochlorite (BLEACH-RITE®, Current Technologies), 0.3 mM hydrogen peroxide (Macron Fine Chemicals), 1.0 ng/µL tetracycline (Sigma-Aldrich®), or 10.0 ng/µL rifampicin (Sigma-Aldrich®) respectively. Bleach stress was increased to 74.6 mM and 149.2 mM on days two and three respectively, and peroxide stress was increased to 0.6 mM on day two and three to maintain selection pressure. Cultures for subsequent experimental days were created as described above and inoculated with 2 µL of the previous day's culture. Culture growth was monitored in 384 microplate wells in a GENios plate reader (Tecan Group Ltd.) operating under the Magellan™ software (version 7.2).

Optical densities were measured at 562 nm absorbance in 20 minute intervals. Temperature was maintained at 37°C, and cultures were shaken for 16.6 minutes after each measurement with an additional 10 seconds of shaking before measurement. Data output was used to construct raw growth curves over multiple days (Supplementary Fig S3), and growth rates (µ) and lag times (τ) were determined using the *GrowthRates* version 1.8[161] and calculated for each day (Extended Dataset). This program estimates the period of exponential growth and excludes lag and stationary phases from calculation

of μ. Data were normalized to the performance of strain MG1655-Control over the course of three days. After normalization, lag times were inverted to simplify analysis for the reader. As larger lag times are considered detrimental, inverting these values made values below 1.0 appear detrimental, as they are for growth rates. Distance from μ_{norm} = 1.0, τ^{-1}_{norm} = 1.0 on each graph was calculated as $D = \sum_n \sqrt{(\mu_n\text{-}1)^2 + (\tau^{-1}{}_n\text{-}1)^2}$ $D = \sum_n \sqrt{(\mu_n\text{-}1)^2 + (\tau^{-1}{}_n\text{-}1)^2}$, where *n* is each of the seven inhibition or activation perturbations strains from their respective graphs.

At the end of adaptation experiments, glycerol stocks of cultures were saved. For re-adaptation experiments, glycerol stocks of the original biological triplicates were streaked onto LB agar plates containing antibiotics and grown overnight. Individual colonies were used to inoculate LB cultures containing antibiotics and grown for 16 hours. Afterwards, cultures were diluted 1/10 into fresh LB containing antibiotics and grown for 24 hours. From this point, the protocol for the original three-day adaptation experiment was performed for cultures under no-stress, no-stress and aTc induction, and the original stress condition (rifampicin or tetracycline) and aTc induction.

5.5.6 Competitive fitness assays.

Gene perturbation strains MG1655-*mutS*a, MG1655-*dinB*i, MG1655-*acrA*i and MG1655-*mutS*a-*dinB*i, as well as the control strain MG1655-Control, were competed against MG1655-mCherry. mCherry fluorescence measured at 610 nm was used to distinguish the two populations during competition experiments using FACS. To measure the fitness of experimental strains relative to MG1655-mCherry, 2 mL LB (or M9 + 0.4% glucose) cultures supplemented with ampicillin, chloramphenicol and aTc were inoculated

from individual colonies and grown overnight for 16 hours. The cultures were then diluted 1:10 in supplemented LB (or M9 + 0.4% glucose) and grown for two hours. Culture ODs (at 562 nm) were then measured and used to mix the two cultures together at either 1:1 or 3:7 OD ratio as indicated in the figures and text. A total of 200 µL of cultures were mixed, out of which 10 µL was added to 190 µL of each stress or non-stress condition per biological replicate. The remaining volume was used for FACS analysis of pre-experiment starting population distributions (D0). Cultures were grown in 96 well microplates in the GENios plate reader as described previously for one day, and the final cultures were collected for FACS analysis.

Relative fitness was determined using the ratio of Malthusian parameters (*m*) of each experimental strain against *m* of competitor strain MG1655-mCherry, as previously described[227]. Malthusians were calculated as $m = \ln (N_f / N_i)$ where N_f and N_i are the number of final and initial cells in each mixture respectively. Initial and final cell counts were determined from FACS analysis, and adjusted to represent their respective dilutions.

5.5.7 Flow cytometry.

Samples for FACS analysis were washed twice in phosphate-buffered saline (PBS) and resuspended in PBS + 4.0% para-formaldehyde (Fisher Scientific). Samples were diluted 1:10 in PBS and sorted using a CyAn ADP analyzer cytometer. Samples were kept on ice throughout the procedure. From each sample 100,000 cells were counted using a voltage of 920 V in a PE-Texas Red channel, a forward scatter gain of 40, and a side scatter voltage of 550 V for detection of mCherry fluorescence. Cells which fluoresced above an intensity of 20 were determined to be MG1655-mCherry, while those

below 20 were determined to be the experimental strains. FACS data was analyzed using Matlab and Summit software.

5.5.8 Sequencing of experimental strains.

For sequencing of genes directly targeted by CRISPR constructs, glycerol stocks of cultures saved at the end of three days of stress exposure were streaked onto LB agar plates containing antibiotics. Two colonies from each biological replicate were picked and used to perform colony PCR using Phusion high-fidelity DNA polymerase. Primers used for these reactions are listed in Supplementary Table S4, and resulted in the fragments listed. PCR products were subsequently gel-purified. Sequencing reactions were performed using each reverse primer by GENEWIZ. Sequences were aligned to *E. coli* MG1655 NCBI reference genome NC_000913 using NCBI BLASTN.

5.5.9 Mutation Fluctuation Assay.

Overall mutation rates of a subset of strains were determined using a mutation fluctuation assay as outlined by Luria and Delbrück[175]. Briefly, single colonies of each strain were picked from LB agar plates supplemented with chloramphenicol and ampicillin and grown overnight for 16 hours in 1 mL LB without selection. These cultures were then adjusted to the same optical densities via addition of LB to denser cultures, and diluted 1:10,000 into thirty three parallel 100 µL cultures of LB supplemented with chloramphenicol, ampicillin and aTc. After 24 hours of growth, three cultures were plated in dilutions on LB agar to determine colony forming units. The remaining 30 cultures were plated on LB agar plates supplemented with 100 µg/mL of rifampicin. Colonies on each

plate were counted after 24 hours of growth in 37°C and used to determine mutation rates. Rates were calculated using the FALCOR web tool[177].

5.5.10 Determination of epistasis.

Expected growth rates were calculated assuming a multiplicative model[268] of deviations from the relative control values: for example, $\mu_{\text{Expected,gene12}} = \mu_{\text{gene1}}{}^{*}\mu_{\text{gene2}}\mu_{\text{Expected,gene12}} = \mu_{\text{gene1}}{}^{*}\mu_{\text{gene2}}$ where $\mu_{\text{gene1}}\mu_{\text{gene1}}$ represents the growth rate observed in the strain targeting gene "11" individually. This provided the expected shift in growth rates (with respect to MG1655-Control) of the multi-gene targeting strains based on results from their individual gene targeting strains. Epistasis when combining perturbation of gene "*a*" and "*b*" was calculated as the difference between the observed relative growth rate of the multi-target strains and the calculated expected growth $\text{Epistasis}_{\text{ab}} = \mu_{\text{observed,ab}} - \mu_{\text{expected,ab}}$ where $\mu_{\text{expected,ab}} = \mu_{\text{a}} * \mu_{\text{b}}$). Calculations of epistasis in inverse lag times were analogous to this procedure. Epistasis was calculated for all five strains with simultaneously perturbed gene expression under all six growth conditions based on the average values presented in Extended Dataset. The 95% confidence interval for average fitness epistasis was calculated using standard error. A z-test was performed to calculate the probability that this deviated from the null hypothesis of no epistasis, and the resulting *P*-value was obtained assuming a two-tailed distribution.

5.6 Author Contributions

P.B.O. performed all experiments. K.E.E. and P.B.O. designed sgRNA constructs, and P.B.O generated dCas9-ω and sgRNA constructs and strains. A.E.B. collected and

analyzed FACS data. K.E.E. designed RT-qPCR sequencing primers. A.C. and P.B.O. conceived of experiments, and P.B.O. and K.E.E. designed experiments. P.B.O. and A.C. wrote the manuscript.

5.7 Supplementary Information

5.7.1 Supplementary Tables

Supplementary Table 5.1 Genes investigated in this study

Major functions for each protein are listed as they are relevant to this study. Note that protein promiscuity means that each might have secondary functions not listed. Information was obtained from ecocyc.org.

Gene	Protein	Function
mutS	MutS	Combines with MutH and MutL to form the MutHLS complex, which is directed by methylation to repair DNA-DNA mismatches. MutS binds to mismatched DNA and directs MutH to cleave the unmethylated strand, allowing for other enzymes to repair the mismatch. In this way, MutS serves to maintain the genomes' status quo.
soxS	SoxS	Dual transcriptional regulator of the superoxide stress response. Binds to common recognition motifs to regulate expression of genes involved in the superoxide regulon. SoxS and MarA share 49% homology, and bind to similar DNA elements such as the Mar-Sox-Rob box. SoxS also regulates expression of some genes controlled only by the Sox box independent of MarA
tolC	TolC	Part of multiple multidrug efflux pump systems, including TolC-AcrAB. The TolC trimeric protein acts as an outer membrane porin to shuttle hydrophobic and amphipathic molecules outside of the cell. TolC also binds to the periplasmic component of AcrAB
acrA	AcrA	Part of the TolC-AcrAB multidrug efflux pump. AcrA is a periplasmic protein which complexs with the inner membrane protein AcrB. It acts as a secondary transporter of molecules from AcrB to TolC as they are shuttled outside the cell.
recA	RecA	Induces the SOS response by cleaving LexA dimers. These dimers bind to the SOS box to constitutively inhibit genes involved in the SOS response. DNA damage (and double strand breaks in particular) activates RecA to cleave LexA, thus freeing repression of the SOS response.
dinB	DNA polymerase IV	Polymerase which lacks proofreading capacity, making it more prone to introducing errors during replication and thus creates spontaneous mutations. The polymerase has particular affinity towards misalignments and DNA lesions.
marA	MarA	Multiple Antibiotic Resistance protein which acts as a dual transcriptional regulator of at least 60 other genes which play roles in protection against antimicrobial stressors. Shares homology with SoxS, and binds to similar DNA elements such as the Mar-Sox-Rob box. MarA also regulates expression of some genes controlled only by the Mar box independent of SoxS.
RFP	Red Fluorescent Protein (RFP)	Produces the Red Fluorescent Protein. This gene was not present in any of the strains used in this study, and thus sgRNAs targeting *RFP* served as a control.
mCherry	mCherry Fluorescent Protein	Produces the molecule mCherry, which also fluoresces red. This gene was present only in MG1655-mCherry. While similar in function to RFP, *mCherry* shares no homology with the sgRNA targeting RFP and thus was not affected by gene perturbation constructs. Homology was determined from the Basic Local Alignment Search Tool, available from NCBI.

Supplementary Table 5.2 Plasmids used in this study

Name	Purpose	Reference or Source
pdCas9	dCas9	Addgene Plasmid 44249 (Qi et al. 2013)
sgRNA-RFP	sgRNA targeting RFP (control)	Addgene Plasmid 44251 (Qi et al. 2013)
pWJ66	dCas9-ω	Addgene Plasmid 46570 (Bikard et al. 2013)
pPO-dCas9ω	dCas9-ω in 44249 vector	This study
pPO-mCherry	sgRNA targeting RFP (control) expressing aTc inducible mCherry	This study
pPO-*mutS*i	sgRNA inhibiting *mutS*	This study
pPO-*soxS*i	sgRNA inhibiting *soxS*	This study
pPO-*tolC*i	sgRNA inhibiting *tolC*	This study
pPO-*acrA*i	sgRNA inhibiting *acrA*	This study
pPO-*recA*i	sgRNA inhibiting *recA*	This study
pPO-*dinB*i	sgRNA inhibiting *dinB*	This study
pPO-*marA*i	sgRNA inhibiting *marA*	This study
pPO-*mutS*a	sgRNA activating *mutS*	This study
pPO-*soxS*a	sgRNA activating *soxS*	This study
pPO-*tolC*a	sgRNA activating *tolC*	This study
pPO-*acrA*a	sgRNA activating *acrA*	This study
pPO-*recA*a	sgRNA activating *recA*	This study
pPO-*dinB*a	sgRNA activating *dinB*	This study
pPO-*marA*a	sgRNA activating *marA*	This study
pPO-*tolC*i-*acrA*i	sgRNA inhibiting *tolC* and *acrA*	This study
pPO-*tolC*i-*acrA*i-*soxS*i	sgRNA inhibiting *tolC, acrA* and *soxS*	This study
pPO-*recA*a-*dinB*a	sgRNA activating *recA* and *dinB*	This study
pPO-*mutS*a-*dinB*i	sgRNA activating *mutS* and inhibiting *dinB*	This study
pPO-*tolC*i-*dinB*i	sgRNA inhibiting *tolC* and *dinB*	This study

Supplementary Table 5.3 Experimental *E. coli* strains used in this study

Name	Plasmids Harbored
MG1655	None
MG1655-i	pdCas9
MG1655-a	pPO-dCas9ω
MG1655-*rfp*i	pdCas9 + sgRNA-RFP
MG1655-*rfp*a	pPO-dCas9ω + sgRNA-RFP
MG1655-Control	sgRNA-RFP + pdCas9 or pPO-dCas9ω
MG1655-mCherry	pPO-sgRNA-mCherry + pdCas9
MG1655-*mutS*i	pdCas9 + pPO-*mutS*i
MG1655-*mutS*a	pPO-dCas9ω + pPO-*mutS*a
MG1655-*soxS*i	pdCas9 + pPO-*soxS*i
MG1655-*soxS*a	pPO-dCas9ω + pPO-*soxS*a
MG1655-*tolC*i	pdCas9 + pPO-*tolC*i
MG1655-*tolC*a	pPO-dCas9ω + pPO-*tolC*a
MG1655-*acrA*i	pdCas9 + pPO-*acrA*i
MG1655-*acrA*a	pPO-dCas9ω + pPO-*acrA*a
MG1655-*recA*i	pdCas9 + pPO-*recA*i
MG1655-*recA*a	pPO-dCas9ω + pPO-*recA*a
MG1655-*dinB*i	pdCas9 + pPO-*dinB*i
MG1655-*dinB*a	pPO-dCas9ω + pPO-*dinB*a
MG1655-*marA*i	pdCas9 + pPO-*marA*i
MG1655-*marA*a	pPO-dCas9ω + pPO-*marA*a
MG1655-*tolC*i-*acrA*i	pdCas9 + pPO-*tolC*i-*acrA*i
MG1655-*tolC*i-*acrA*i-*soxS*i	pdCas9 + pPO-*tolC*i-*acrA*i-*soxS*i
MG1655-*recA*a-*dinB*a	pPO-dCas9ω + pPO-*recA*a-*dinB*a
MG1655-*mutS*a-*dinB*i	pPO-dCas9ω + pPO-*mutS*a-*dinB*i
MG1655-*tolC*i-*dinB*i	pdCas9 + pPO-*tolC*i-*dinB*i

Supplementary Table 5.4 Cloning and sequencing primers utilized in this study

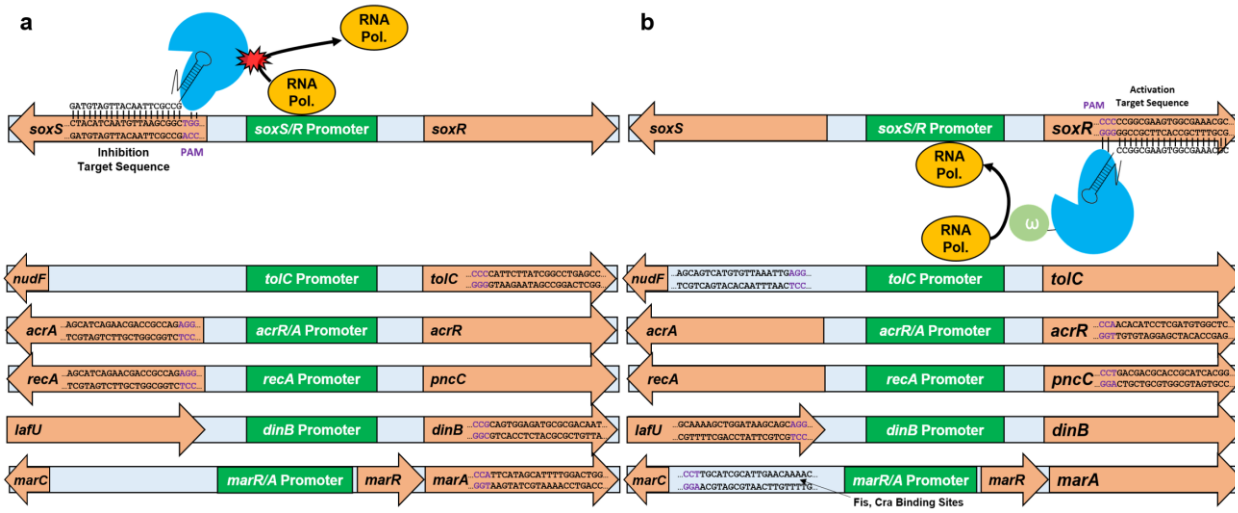
Note that reverse *recA* sequencing 1 was used to synthesize the full *recA* sequencing product. Both sequencing reactions were performed with both reverse *recA* sequencing 1 and 2.

Name	Sequence (5'-3')
Forward *mutS*-a	ACTAGTACTAGTGCAAGTACGCAAAATTGTATGTTTTAGAGCTAGAAATAGC
Forward *mutS*-i	ACTAGTACTAGTCTGCTGCATCATGGGCGTATGTTTTAGAGCTAGAAATAGC
Forward *soxS*-a	ACTAGTACTAGTGCGTTTCGCCACTTCGCCGGGTTTTAGAGCTAGAAATAGC
Forward *soxS*-i	ACTAGTACTAGTCTACATCAATGTTAAGCGGCGTTTTAGAGCTAGAAATAGC
Forward *tolC*-a	ACTAGTACTAGTAGCAGTCATGTGTTAAATTGGTTTTAGAGCTAGAAATAGC
Forward *tolC*-i	ACTAGTACTAGTGGCTCAGGCCGATAAGAATGGTTTTAGAGCTAGAAATAGC
Forward *acrA*-a	ACTAGTACTAGTGAGCCACATCGAGGATGTGTGTTTTAGAGCTAGAAATAGC
Forward *acrA*-i	ACTAGTACTAGTAGCATCAGAACGACCGCCAGGTTTTAGAGCTAGAAATAGC
Forward *recA*-a	ACTAGTACTAGTCCGTGATGCGGTGCGTCGTCGTTTTAGAGCTAGAAATAGC
Forward *recA*-i	ACTAGTACTAGTTACCAAATTGTTTCTCAATCGTTTTAGAGCTAGAAATAGC
Forward *dinB*-a	ACTAGTACTAGTGCAAAAGCTGGATAAGCAGCGTTTTAGAGCTAGAAATAGC
Forward *dinB*-i	ACTAGTACTAGTATTGTCGCGCATCTCCACTGGTTTTAGAGCTAGAAATAGC
Forward *marA*-a	ACTAGTACTAGTGTTTTGTTCAATGCGATGCAGTTTTAGAGCTAGAAATAGC
Forward *marA*-i	ACTAGTACTAGTCCAGTCCAAAATGCTATGAAGTTTTAGAGCTAGAAATAGC
Forward dCas9-ω	AGATCTAGATCTAAAGAGGAGAAAGGATCTATGGATAAGAAATACTCAAT
Reverse dCas9-ω	CTCGAGCTCGAGTTAACGACGACCTTCAGCAA
sgRNA sequencing	GGGGGGGACGTCTAAGAAACCATTATTATCATG
Reverse *sgRNA*	GGGCCCGGGCCCAAGCTTCAAAAAAAGCACCG
Forward *soxS* sequencing	CTATTGCCAGGGATGGTTC
Reverse *soxS* sequencing	TTTCATAGAAATGCAGCGCC
Forward *recA* sequencing	GGATGTTGATTCTGTCATGG
Reverse *recA* sequencing 1	TATGCATTGCAGACCTTGTG
Reverse *recA* sequencing 2	AGTAGACGTTATCGTCGTTG

Supplementary Table 5.5 RT-qPCR primers utilized in this study

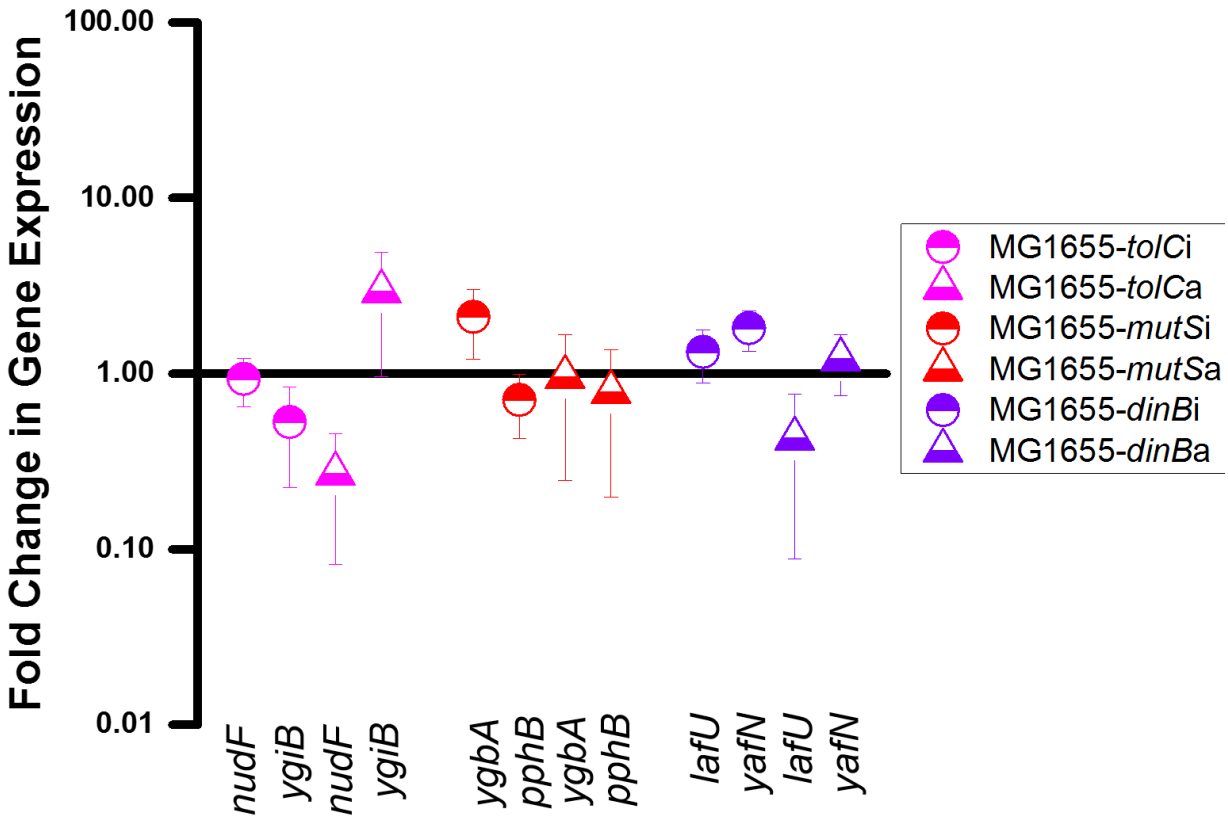
Name	Sequence (5'- 3')
mutS forward	ATGGAACGTGAGCAGGACAG
mutS reverse	CAGCCAGCGTTTCAGCATAC
soxS forward	TCTGCTGCGAGACATAACCC
soxS reverse	ACTTGCAACGAATGTTCCGC
tolC forward	ACGCACTACCACCAGTAACG
tolC reverse	TTTGTCTTCCGGGACCAGTG
acrA forward	AAGCCCTTCTTCCAGACGTG
acrA reverse	AACGGCAAAGCCAAAGTGTC
recA forward	ATCGCCTGGCTCATCATACG
recA reverse	GCACTGGAAATCTGTGACGC
dinB forward	GGCCAGTTTGTGATTACGCC
dinB reverse	CTACGCTCCCACAAAATGCG
marA forward	AATCGCGCAAAAGCTGAAGG
marA reverse	GCGATTCGCCCTGCATATTG
rrsA forward	AACACATGCAAGTCGAACGG
rrsA reverse	AATCCCATCTGGGCACATCC
nudF forward	CGCAGTCTTGCTACCCTTTG
nudF reverse	ACGGGCAACATCTTCCACAC
ygbA forward	GCAAGCGTATCTCTCGTGAA
ygbA reverse	CGCCGAACACACATTTATCC
lafU forward	TCGGACGCACTTTTGGTCAG
lafU reverse	CAACTGGAACCTTTCGGGTG
yafN forward	TGATCAACCGGTTGCGGTTC
yafN reverse	ATCTTGCAGCACTTGGACGG
ygiB forward	AATACGCCACCCGTGAAGAC
ygiB reverse	GACGCCCCATCATGTAACCG
pphB forward	GGAGAGCGAATTACTCTGGC
pphB reverse	GGTTAGCGAACGTCTGAATG

5.7.2 Supplementary Figures



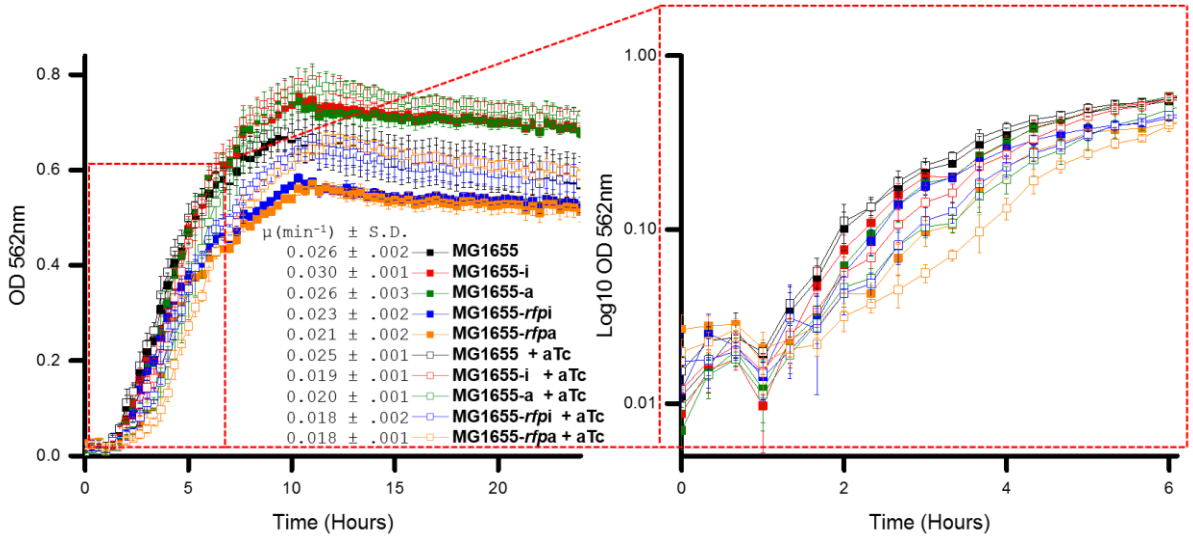
Supplementary Figure 5.1 Genetic context of CRISPR inhibition and activation targets

Target sequence of single-target sgRNA constructs. These targets are analogous to Figure 1B. (A) Inhibition target sequences, in which a demonstration of construct-induced gene inhibition (via exclusion of RNA polymerase) is shown for soxS. Corresponding inhibition target sequences are shown for tolC, acrA, recA, dinB and marA. (B) Activation target sequences, in which a demonstration of construct-induced gene activation (via recruitment of RNA polymerase) is shown for soxS. Corresponding activation target sequences are shown for tolC, acrA, recA, dinB and marA. Potential genomic regions of interest are also included. Downstream genes within the same operon include: acrB for acrA, ygiB-D for tolC, recX for recA, yafN-P for dinB and marR/B for marA. The genes mutS and soxS have no other known genes within their operon, although the pphB start site is located 105 nt downstream of mutS indicating potential transcriptional overlap. Aside from marA's feedback network mentioned in the text, no compounding regulatory interactions have been reported within each operon.



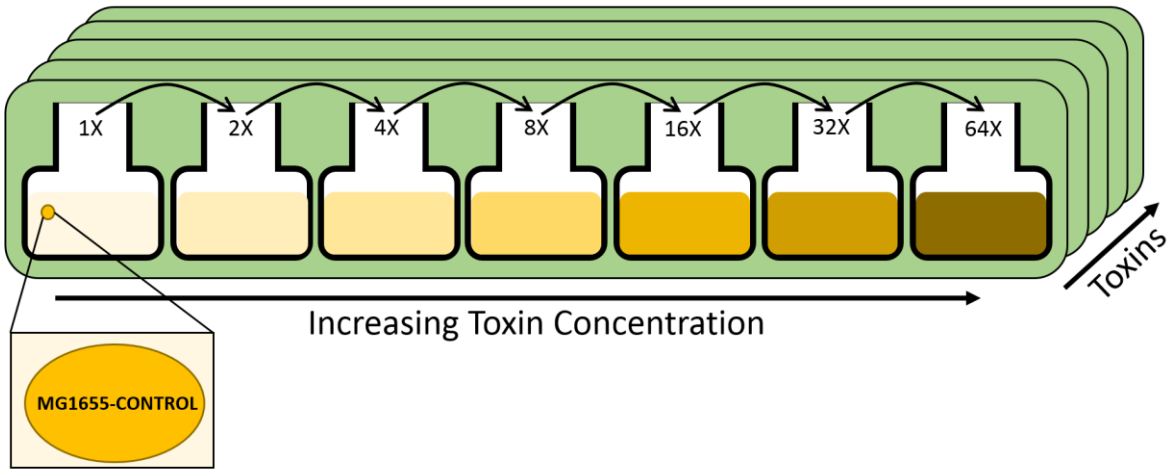
Supplementary Figure 5.2 Impacts of perturbation constructs on neighboring genes

qPCR analysis was performed on genes upstream (*nudF*, *ygbA*, *lafU*) and downstream (*ygiB*, *pphB*, *yafN*) of the intended CRISPR perturbation targets *tolC*, *mutS* and *dinB*. Relative fold changes in gene expression from the wild type were quantified using the same approach outlined in Figure 1D. Minor off-target effects were expected in some cases; for instance, the *dinB* activation target overlapped with the 3' end of the *lafU* ORF. Consequently, MG1655-*dinB*a exhibited decreased *lafU* expression (0.43 ± 0.34) with respect to the control MG1655 strain. Similarly, the activation target for *tolC* overlapped with the promoter of *nudF*, making its decrease in expression in MG1655-*tolC*a (0.27 ± 0.19) expected. Inhibition and activation of *tolC* caused decreased and increased expression respectively of the downstream gene *ygiB* by 0.53 ± 0.31 and 2.93 ± 1.97 respectively. The slight increase of both *ygbA* in MG1655-*mutS*i (2.10 ± 0.90) and *yafN* in MG1655-*dinB*i (1.81 ± 0.47) were unexpected, and could indicate potential regulatory overlap of these gene's transcriptional regions (both *ygbA* and *yafN* have yet to be fully characterized). Negligible effects were observed in the remainder of the samples.



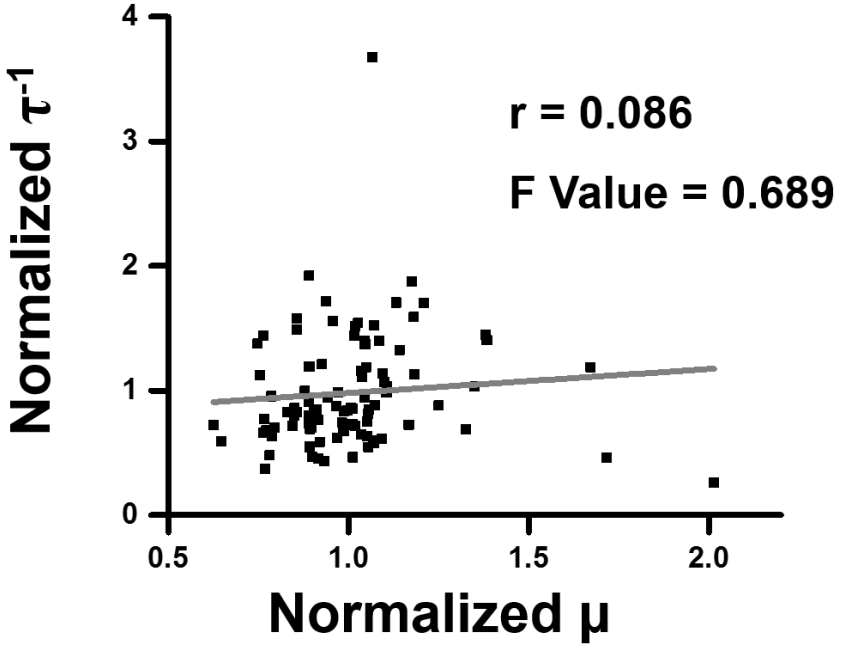
Supplementary Figure 5.3 Growth of MG1655 strains carrying sgRNA and dCas9 (or dCas9-ω) constructs confirms cell viability

Individual dCas9 (MG1655-i) and dCas9-ω (MG1655-a) plasmids were transformed into MG1655 cells, and were also transformed along with *rfp*-targeting sgRNA controls (MG1655-*rfp*i and MG1655-*rfp*a). Cultures were grown overnight either in the absence of aTc induction (solid lines) or in the presence of aTc induction (dashed curves). aTc was used to induce expression of dCas9 or dCas9-ω. Growth rates of each strain (presented left of the legend) were calculated using *GrowthRates* version 1.8, which excludes the estimated lag phase. As expected, a slight decrease in growth was observed for MG1655 strains carrying two plasmids (blue and orange lines), as well as during aTc induction. However, under similar conditions, dCas9 and dCas9-ω plasmids induced no discernable differences in growth rates. A two-tailed t-test showed no statistically significant differences between the MG1655-i and MG1655-a in the absence (P = 0.17) and presence (P = 0.36) of aTc. Similarly, no statistically significant difference was found between MG1655-*rfp*i and MG1655-*rfp*a in the absence (P = 0.33) and presence (P = 0.90) of aTc. An average growth rate of 0.018 min^{-1} of MG1655-*rfp*a strain (which was used as the MG1655-Control strain) corresponds to 38.5 min of doubling time, which gives rise to approximately 7.9 generations during the 5 hours of exponential growth (right panel, magnified from left panel and replotted on a log-scale). Error bars (and s.d.) represent s.d. of biological triplicates.



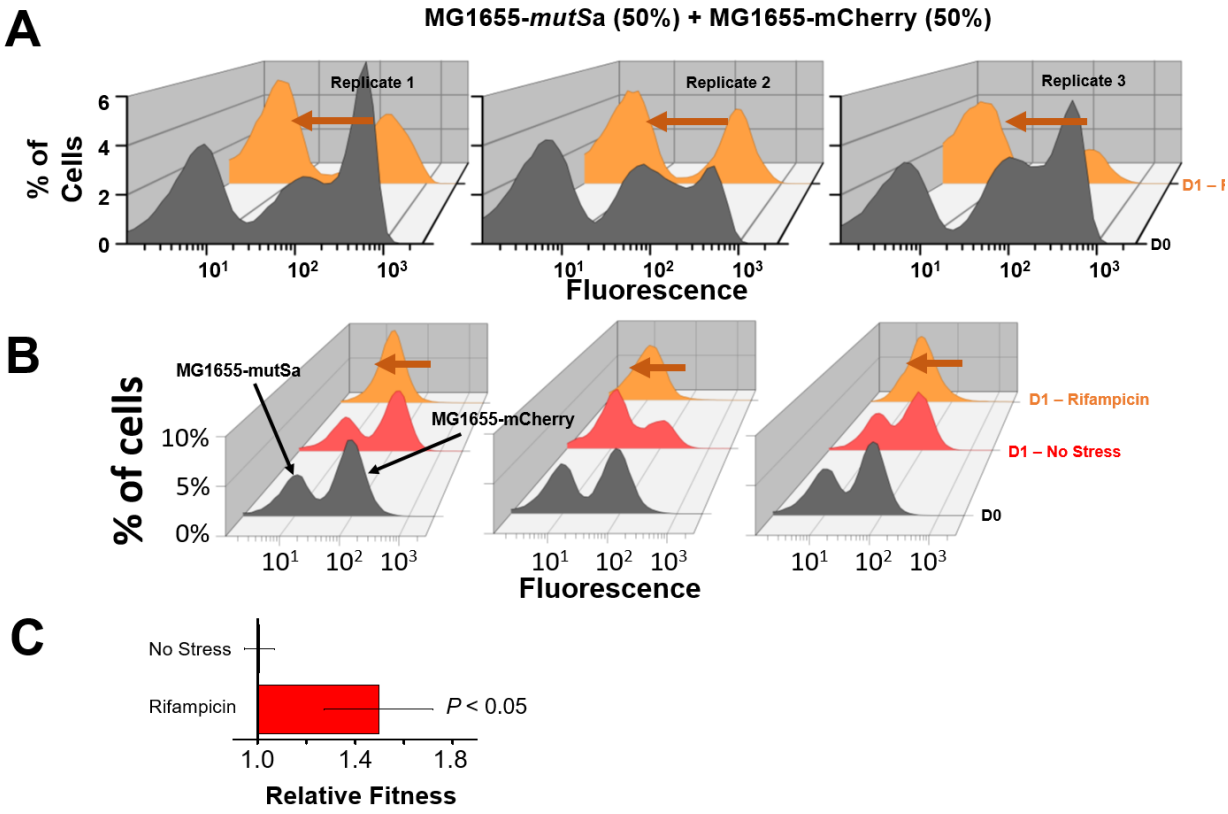
Supplementary Figure 5.4 Experimental design for determination of Minimum Inhibitory Concentrations (MICs)

Before beginning adaptation experiments, the MIC of each toxin or nutrient was determined using the MG1655-*rfp*i strain harboring the RFP-targeting sgRNA control plasmid, along with the dCas9 plasmid induced with 10 ng/μL aTc. A 2-fold serial dilution in concentration was used to test seven concentrations for each toxic compound used in the final adaptation experiments shown in Figure 2 and Figure 3. The MIC was determined to be the lowest concentration that prevented a 0.1 increase in OD_{562nm} after overnight growth in each stress.



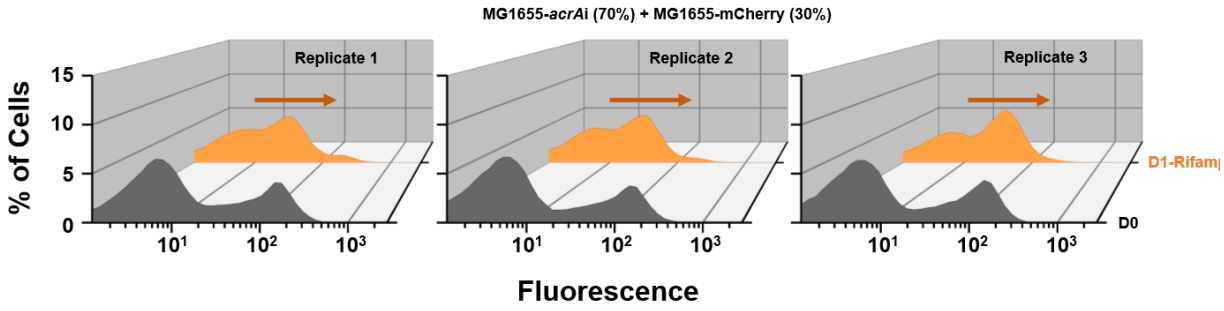
Supplementary Figure 5.5 Linear fits of growth characteristics

Normalized τ^{-1} and μ from all 14 single-gene perturbation strains grown under the five stress conditions were used to develop a linear fits between the two growth characteristics. Each data point correlates to two of these growth characteristics of a strain grown under one condition (bleach, peroxide, tetracycline, rifampicin and nutrient limitation stress). A total of 84 data points were used for each linear fit. Pearson's product-moment correlation coefficient (r), as well as the F value determined from ANOVA is presented. As these values indicate that the slope is not statistically different than 0 at the *P*-value ≤ 0.05 level, the variables can be considered independent from one another. This means that the growth rates and lag times presented in this study each provided unique phenotypic insights decoupled from one another.



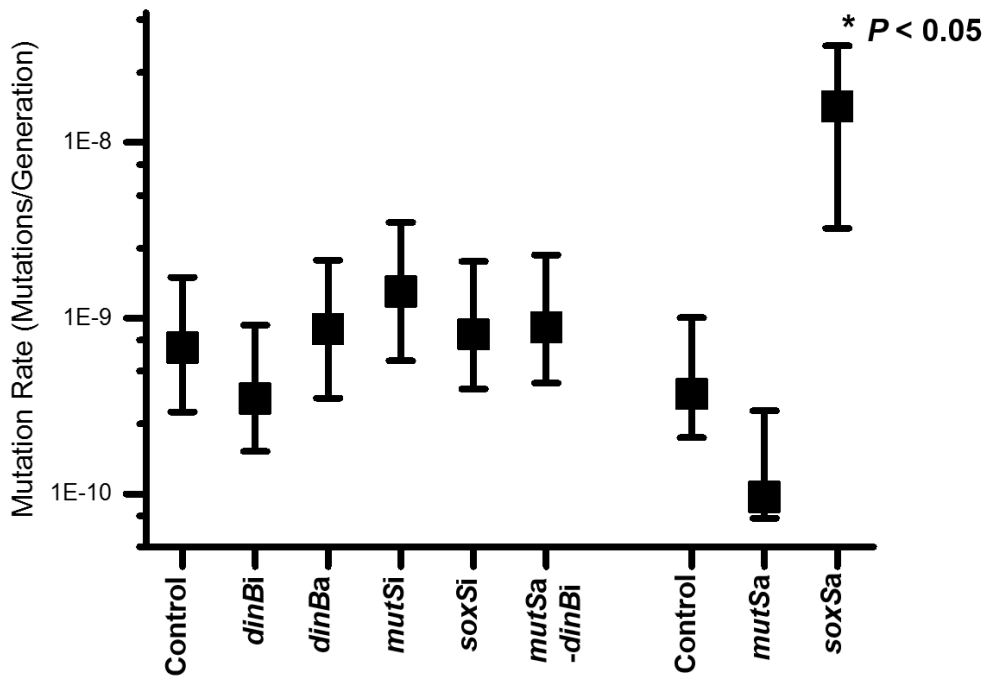
Supplementary Figure 5.6 Variations of competition of MG1655-*mutS*a and MG1655-mCherry

Utilizing the same approach outlined in Fig 3D, MG1655-*mutS*a was competed against MG1655-mCherry in equal concentrations (50% by OD) in both (**A**) M9 minimal media + 0.4% glucose in lieu of LB (total cell count was reduced from 100,000 to 25,000 under these conditions to account for lower culture density), as well as (**B**) LB. Both conditions were used to demonstrate that the phenomenon presented in Fig 3D was dependent on neither the starting concentration of cells, nor the choice of media used. This is demonstrated by a shift in the distribution of cells towards the non-fluorescent MG1655-*mutS*a in both media during rifampicin exposure. In M9, the relative starting percentage of MG1655-*mut*Sa was quantified using FACS to be 37%, 49%, and 36% for replicates 1-3 respectively, and the relative ending percentage after 1 day of rifampicin exposure was 66%, 67%, and 81% (calculated as fraction of cells demonstrating fluorescence below the threshold of 20 during FACS). In LB, the relative starting percentage of MG1655-*mutS*a was 33%, 42%, and 39% for replicates 1-3 respectively, and the relative ending percentage was 32%, 54%, and 32% after one day of no stress exposure, and 46%, 77%, and 71% after one day of rifampicin exposure. (**C**) Relative fitness levels of MG1655-*mutS*a in relation to MG1655-mCherry after one day of exposure to either no stress or rifampicin stress, using data from (**B**). MG1655-*mutS*a fitness under rifampicin stress was significantly greater than its same fitness under no stress exposure. *P*-values were calculated using a two-tailed type II *t*-test, and error bars represent s.d. of biological triplicates.



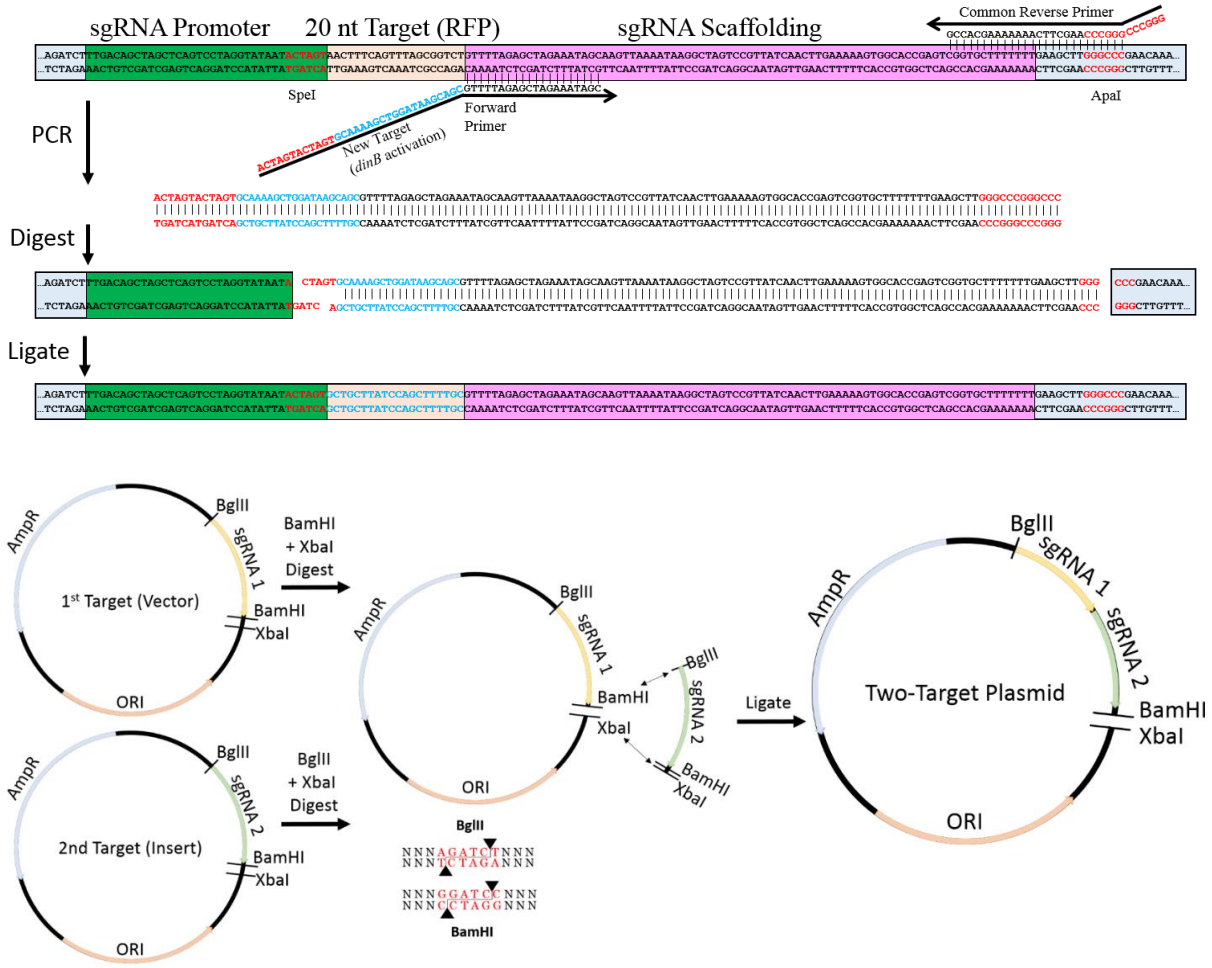
Supplementary Figure 5.7 Competition of MG1655-*acrA*i and MG1655-mCherry

Utilizing the same approach outlined in Figure 3B-E, MG1655-*acrA*i was competed against MG1655-mCherry. A higher starting concentration (70% by OD) of the former was utilized to demonstrate that the MG1655-*acrA*i population was selected against under rifampicin stress, and therefore was less fit.



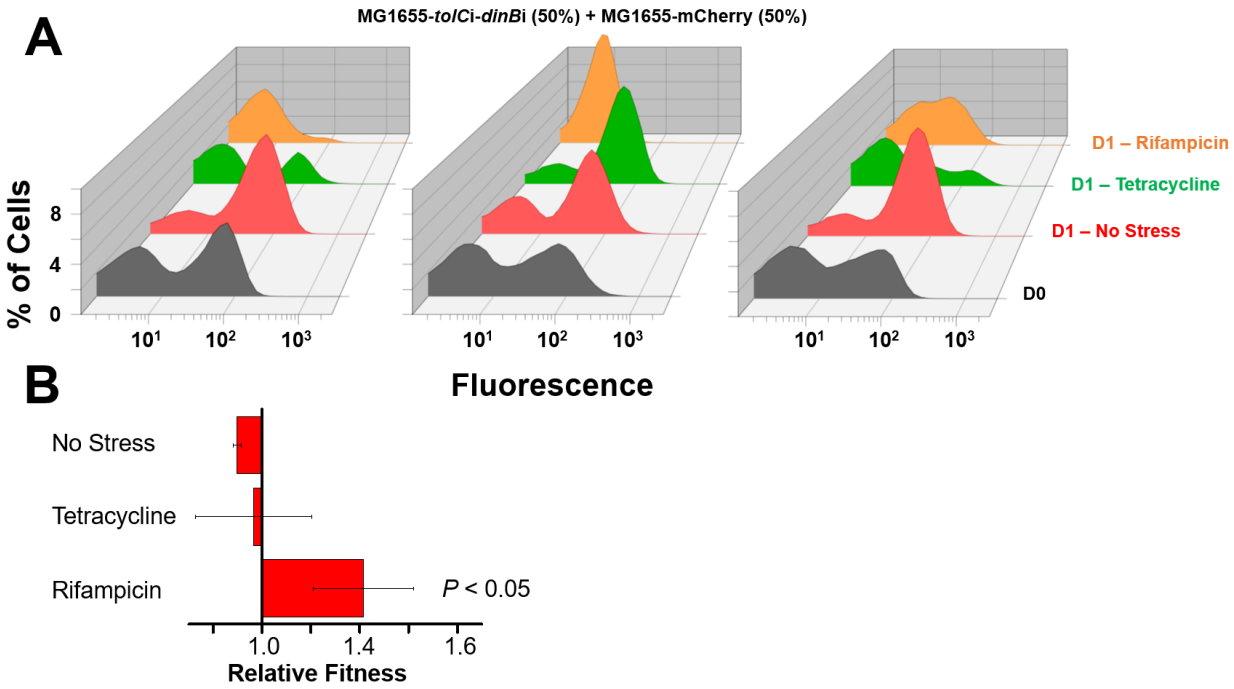
Supplementary Figure 5.8 Mutation fluctuation assay

A subset of experimental strains were subjected to fluctuation tests on rifampicin as described in the Methods. Error bars represent limits of 95% confidence interval. X-axis labels indicate various MG1655 strains, in which the later three were performed in a separate round of experiments. Mutation rates of all strains besides MG1655-*soxS*a were not significantly different than MG1655-Control, although the distribution was shifted towards directions that might, in general, be predicted of particular perturbations. This includes a skew towards lower mutation rates for MG1655-*dinB* and MG1655-*mutS*a, as well as a skew towards higher mutation rates in MG1655-*dinB*a and MG1655-*mutS*i. Interestingly, MG1655-*soxS*a mutation rates were found to be significantly increased over MG1655-Control (P<0.05), indicating underlying complexity in the relationship between gene perturbations and mutation rates.



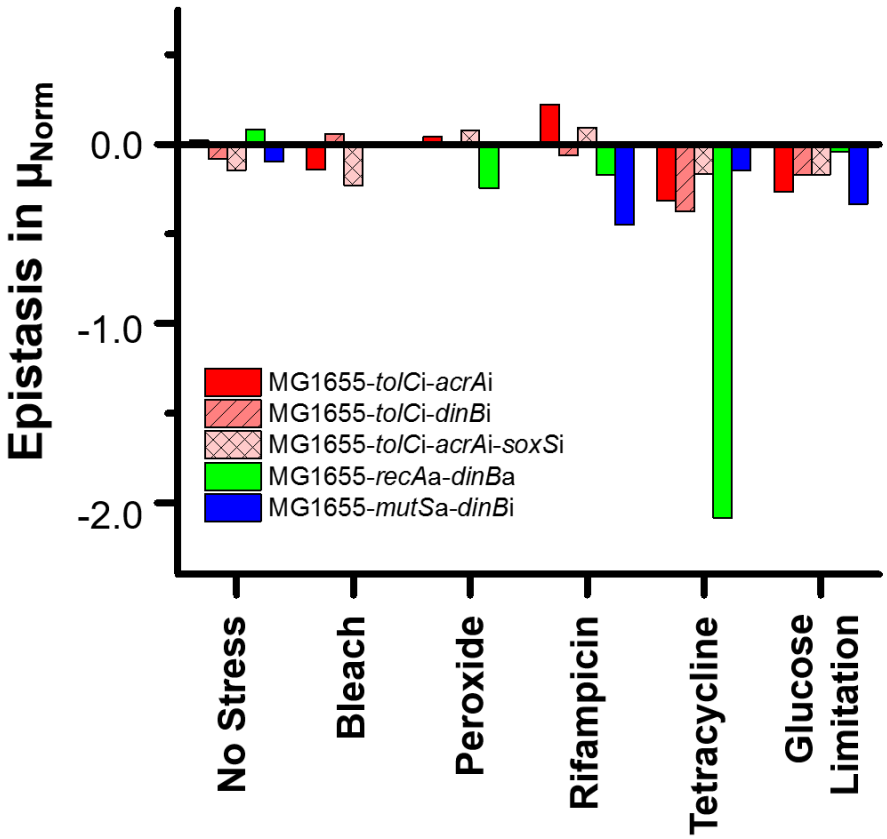
Supplementary Figure 5.9 Experimental design for constructing single and multi-target plasmids

(Top) Cloning of single target sgRNA plasmids. A unique forward primer coding for the new 20 nt target, as well as overlap with the sgRNA scaffolding, was used along with a common reverse primer to create new sgRNA target inserts using PCR. These PCR products, along with the sgRNA plasmid, were digested and ligated to form new sgRNAs. Sequencing was used to confirm each sgRNA target after construction. (Bottom) Cloning of multi-target sgRNA plasmids. The unique sgRNA target "2" is cloned into the plasmid containing sgRNA target "1". Both plasmids are digested with XbaI, while the first and second plasmids are digested with BamHI and BglII respectively. Due to the compatible sticky ends that are generated, the second sgRNA can be ligated into the first plasmid without regenerating a restriction site. This process can be repeated to create sgRNA plasmids targeting "n" number of genes.



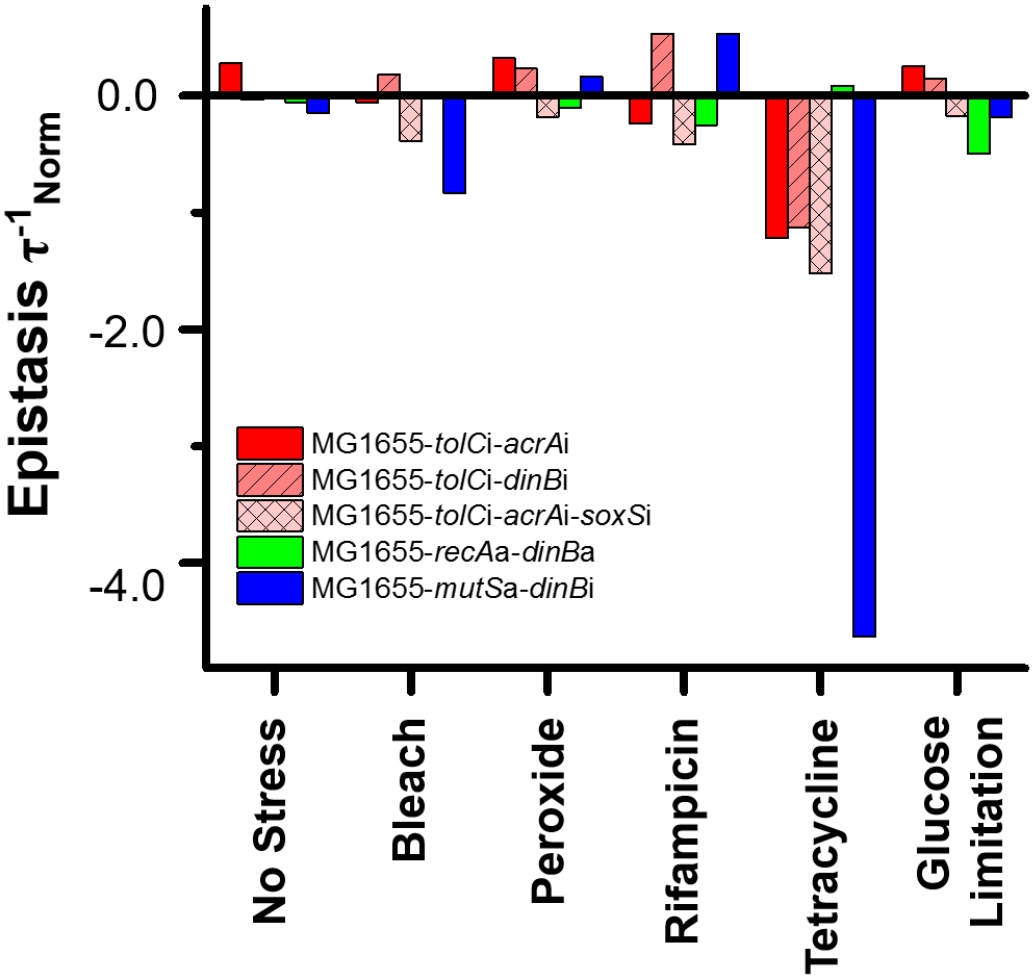
Supplementary Figure 5.10 Competition of MG1655-*tolC*i-*dinB*i and MG1655-mCherry

Utilizing the same approach outlined in Fig 5E, MG1655-*tolC*i-*dinB*i was competed against MG1655-mCherry in equal concentrations (50% by OD) in LB. (**A)** Fluorescence distribution before and after one day of stress (or no stress) exposure. This competition data corroborates the phenotypic results presented n Fig 5C of MG1655-*tolC*i-*dinB*i, showing a shift in the population distribution towards MG1655-*tolC*i-*dinB*i only after rifampicin exposure. (**B**) Malthusian fitness calculation of competition results using the same approach outlined n Fig 5F. Here, we show a significant difference between the fitness of MG1655-*tolC*i-*dinB*i after one day of rifampicin and no stress exposure, but no difference between tetracycline and no stress exposure. *P*-values were calculated using a two-tailed type II *t*-test, and error bars represent s.d. of biological triplicates.



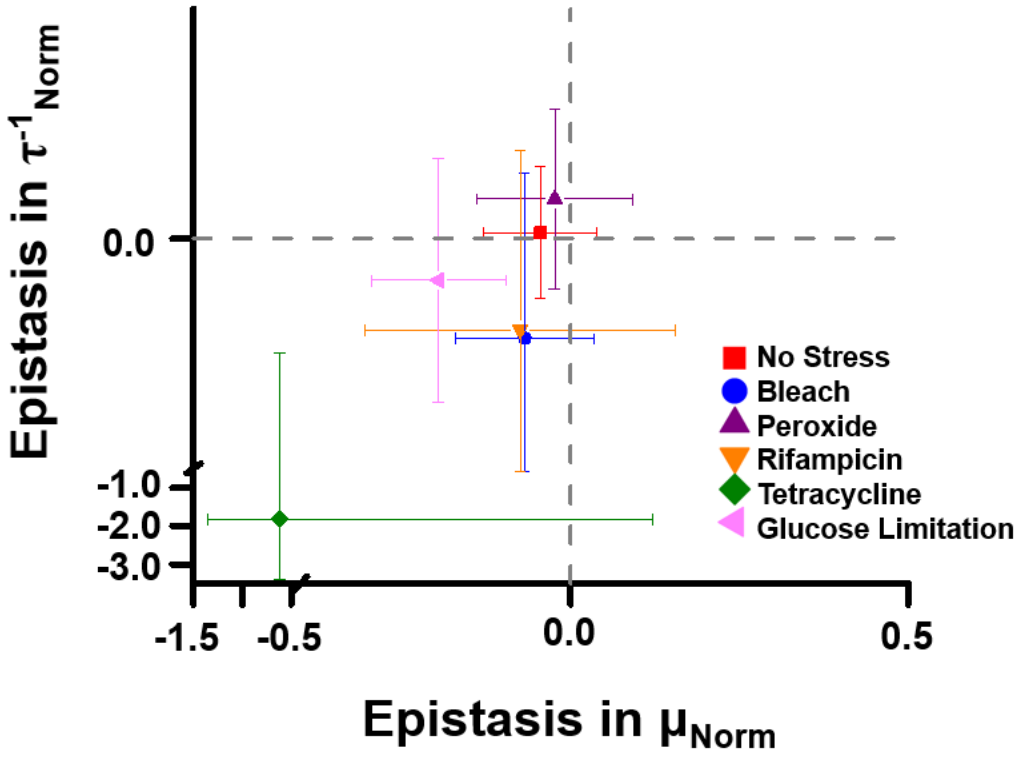
Supplementary Figure 5.11 Epistatic interactions on μ_{norm} of multiple-gene targeting strains

Epistasis was calculated using a multiplicative model based on average μ_{norm} of single target constituents, and of the measured μ_{norm} of multiple-gene targeting strains as presented in the Extended Dataset available online.



Supplementary Figure 5.12 Epistatic interactions on τ^{-1}_{norm} of multiple-gene targeting strains

Epistasis was calculated using a multiplicative model based on average τ^{-1}_{norm} of single target constituents and of the measured τ^{-1}_{norm} of multiple-gene targeting strains as presented in the Extended Dataset available online.



Supplementary Figure 5.13 Distribution of estimated epistatic impacts on τ^{-1}_{norm} and μ_{norm} clustered by stress

Data (and error bars) represent mean (and s.d. of mean) epistasis of the five multiple gene perturbation strains under each stress.

Chapter 6: CHAOS – Deterring bacterial adaptation via epistatic gene expression perturbations

Otoupal, P. B., Cordell, W. T., Bachu, V., Sitton, M. J., & Chatterjee, A. *Submitted.*

6.1 Abstract

The ever-increasing threat of multi-drug resistant bacteria, a shrinking antibiotic pipeline, and the innate ability to adapt necessitates long-term strategies to slow the evolution of antibiotic resistance. Here we develop an approach, dubbed “Controlled Hindrance of Adaptation of OrganismS” or “CHAOS”, involving induction of epistasis between gene perturbations to deter adaption. We construct a combinatorial library of multiplexed, deactivated CRISPR-Cas9 devices to systematically perturb gene expression in *Escherichia coli.* While individual perturbations improved fitness during antibiotic exposure, multiplexed perturbations caused large fitness loss in a significant epistatic fashion. Strains exhibiting epistasis adapted significantly slower or not at all over three to fourteen days, and loss in adaptive potential was shown to be sustainable. Finally, we show that multiplexed peptide nucleic acids increase the antibiotic susceptibility of clinically isolated Carbapenem-resistant *E. coli* in an epistatic fashion. Together, these results suggest a new therapeutic strategy for restricting the evolution of antibiotic resistance.

6.2 Introduction

The rapid emergence of multi-drug resistant “superbugs” poses a serious threat to millions and constitutes an impending international health crisis[14]. Bacteria are constantly

driven to adapt to new treatments, establishing a biological arms race between evolution and our ability to develop new antimicrobial treatments. As the balance of these processes has trended towards more resistance and fewer novel therapies[269], attention must be turned to new strategies which hinder the evolution of antibiotic resistance if the utility of our antibiotic arsenal is to be preserved. Current approaches are largely limited to the cycling of antibiotics, which rely upon high levels of coordination between clinicians and has been found to be difficult to implement in practice[270,271]. We propose an alternative strategy that we dub "Controlled Hindrance of Adaptation of OrganismS" or "CHAOS" that involves introducing multiplexed gene expression perturbations to disrupt bacterial homeostasis by manipulating a fundamental process underpinning evolution: epistasis.

Epistasis describes the nonlinear outcome of combining two or more genetic changes; negative (or positive) epistasis occurs if the combined changes produce a worse (or better) fitness than expected based upon their individual impacts. This phenomenon creates the rugged shapes of fitness landscapes used to describe evolution[249,267,272]. Such landscapes project an organism's fitness at every unique genetic state for a pair of genes, with epistatic interactions influencing the landscape's shape. Epistasis has been found to emerge frequently upon introduction of simultaneous mutations and is widely recognized to influence evolutionary trajectories[273–275]. Negative epistasis in particular has been shown to restrict rates at which bacteria adapt[235,236]. Research regarding epistasis has thus far focused on the theoretical side, using epistasis to explain how particular mutations emerged during evolution. Here we demonstrate that artificial induction of epistasis has the potential for curtailing evolution. This is accomplished by

introducing multiplexed gene expression perturbations into a bacteria that have cascading epistatic effects, restricting its ability to evolve (Fig. 1a).

Bacterial gene expression is stochastic in nature, allowing for heterogeneity throughout a genetically identical population. This provides an evolutionary advantage, in which some subpopulations with altered expression states can survive immediate exposure to a new and stressful environment[111,114,131,276,277]. However, such noise in gene expression comes at a fitness cost, as deviations from basal gene expression are often more deleterious than beneficial[110,231,251,278] (Fig. 1b). This is strongly evidenced by tight expression of essential genes interacting with many genetic partners[17,217,230]. The high potential for deviations in these genes to trigger epistasis has been proposed as an evolutionary force constraining excessive heterogeneity[217]. We hypothesize that this could be exploited as a therapeutic strategy in which accumulating deviations in the expression of multiple genes would cause artificial induction of negative epistasis (Fig. 1c). As the shape of fitness landscapes used to describe evolution are directly related to the degree of epistasis, induction of epistasis can alter the adaptive trajectory available to a bacteria. Multiplexed perturbations could thus constrain an organism from reaching an adapted state, and offers a tangible avenue towards curbing the ability of bacteria to adapt to antibiotic treatment. Previous work lends credibility to this hypothesis[16,19,218].

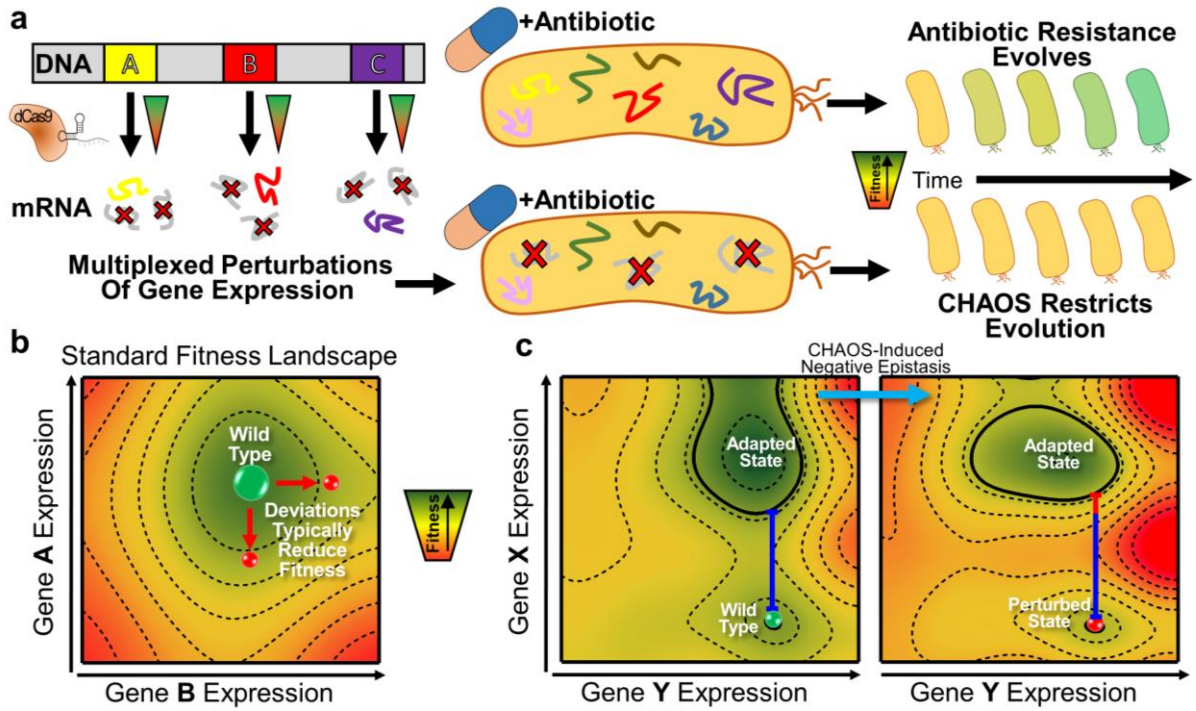


Figure 6.1 Controlled Hindrance of Adaptation of OrganismS (CHAOS) approach to deter evolution

a In CHAOS, basal levels of gene expression are perturbed in a targeted and multiplexed fashion. Introduction of these perturbations alongside an antibiotic disrupts the evolutionary forces driving the organism to evolve, thus restricting the increase in bacterial tolerance to said antibiotic over time. **b** The CHAOS approach is built upon the ideas used to explain how organisms evolve: fitness landscapes and epistasis. Here, a typical two-dimensional fitness landscape is presented based on various expression states of genes A and B (contours represent equal fitness). Although two genes are presented here, such a landscape can be constructed from any combination of an organism's n × n genes. An underlying assumption of these landscapes is that epistasis has molded their shape such that deviations in gene expression from this basal state are typically detrimental to survival. **c** Wright's shifting balance model of evolution[267] predicts that in a small subset of these landscapes there exist adapted states where deviations in gene expression from basal wild-type levels improve fitness. However, these adapted states are separated by regions of fitness lower than the wild-type state that constrains the population. CHAOS can be employed to introduce multiplexed perturbations of gene expression that cause cascading epistatic effects, thereby reshaping the fitness landscape. This epistasis can be tuned to amplify the barrier to reach the adapted state, thus artificially constraining evolution.

Here we explore CHAOS-induced epistasis to restrict bacterial adaptation to antibiotics. We employ deactivated Cas9 (dCas9) Clustered Regularly Interspaced Short

Palindromic Repeats (CRISPR) technology to selectively inhibit[248] or activate[45] mRNA production of specific genes. We construct a library of CRISPR devices and systematically quantify the impact of multiplexed gene perturbations on bacterial fitness with and without antibiotic exposure. We show that increasing the number of genes perturbed from their basal levels strongly correlates with reduced fitness during antibiotic exposure. We prove that this phenomenon arises in a negative epistatic fashion, wherein compounding perturbations exacerbate fitness loss. The degree of this effect correlated with the number of secondary gene partners impacted by the primary gene perturbation(s), with disruption of metabolic processes demonstrating a dominant impact on epistasis. We demonstrate that negative epistasis correlated with significantly diminished adaptation rates over three to fourteen days and that such effects are sustainable, suggesting that multiplexed gene perturbations can be employed to restrict the evolution of antibiotic resistance. Finally, we use peptide nucleic acids (PNAs) to repress expression of a different set of gene perturbations in a multi-drug resistant clinical isolate of *E. coli*, demonstrating that CHAOS is applicable to other modes of gene expression manipulation. We successfully re-sensitized this antibiotic-resistant isolate to treatment, suggesting the benefit of CHAOS in clinical settings.

6.3 Results

6.3.1 Construction of CHAOS gene perturbation library

To explore artificial induction of negative epistasis, we constructed a library of *Escherichia coli* strains harboring dCas9[248] or dCas9-ω[45] devices to inhibit or activate mRNA production respectively. Each strain hosts a unique array of one or more single

guide RNAs (sgRNAs) to direct the CRISPR enzymes to particular genomic loci (Fig. 2a-b).

We designed CHAOS strains targeting two gene sets. The first set included activation of four universal stress response genes important for adaptation[16,19]: *mutS* (DNA mismatch repair), *soxS* (SOX pathway regulator), *tolC* (multidrug efflux pump) and *recA* (SOS response activator) (Fig. 2a). These four genes were determined in our previous work to be particularly important during bacterial adaptation to a broad range of antibiotics and other stressors, and we reasoned that upregulating these genes would serve as the most likely avenue influencing adaptation[16,17,19]. Furthermore, we hypothesized that activating these genes would serve as the "worst-case" scenario, as we expect broad upregulation of stress response to provide multiple avenues of adaptive escape and encourage positive epistasis. The second set included inhibition of four genes found to be universally conserved across a broad set of bacteria central to distinct cellular pathways (Fig. 2b). We began with a set of 174 genes conserved across diverse bacterial genomes[279], from which we selected genes that were not present in operons to minimize the number of genes directly perturbed by each CRISPR construct. From this, we selected well-characterized genes that were involved in diverse cellular processes in order to maximize potential epistatic effects. This led us to four promising conserved gene targets: *dfp* (synthesizes essential coenzyme A), *topA* (an essential supercoiling-relaxing enzyme), *zwf* (a key glycolysis enzyme) and *frr* (essential for ribosome recycling). These genes are monocistronic (except *tolC* and *recA*) and possess little direct interactions between one another (except *mutS/recA* and *soxS/tolC*) (Supplementary Fig. 1).

We analyzed the STRING Database to determine the direct protein-protein interactions known for each gene and gene combination[280]. Multiplexing perturbation of conserved genes was predicted to cause more cascading impacts than multiplexing perturbation of stress response genes (Supplementary Figs. 2-3). We predicted that perturbations within the conserved gene set would thus have a greater impact on fitness, as the cascading effects of the primary gene perturbation would affect more downstream partners in more diverse pathways.

We created CHAOS constructs with all possible single, double, triple and quadruple combinations of gene perturbations for each of these sets (Supplementary Figs. 4-5 and Supplementary Table 1). A two-plasmid system was utilized to induce gene expression perturbation; the first plasmid encoded sgRNA target sequence(s), while the second plasmid encoded an anhydrotetracycline (aTc) inducible dCas9 or dCas9-ω for gene inhibition or activation respectively. Activation sgRNAs targeted ≈80-100 nt upstream of the +1 transcription start site of each gene[45]. Inhibition sgRNAs targeted the +1 transcription start site to inhibit transcriptional read-through via roadblock mechanism[248]. We engineered to cause an approximately 10-fold range of under- or over-expression and verified the degree of perturbation using RT-qPCR (Supplementary Fig. 6). The only unexpected result involved *topA* perturbation, which exhibited a likely Fis-dependent response (explored further in Supplementary Fig. S6b).

We constructed two control strains targeting “nonsense” perturbation of the red fluorescent protein (*rfp*) gene, which was absent in all strains. These strains harbored one or four copies of *rfp* targeting sgRNAs (hereafter referred to as C$_{ontrol}$ and CCCC, respectively). Another control strain including constitutively expressed mCherry and one

copy of *rfp* perturbation ($C_{mcherry}$, see methods) was created to enable tracking of the control population during strain competition. Finally, two control strains harboring either one or four sgRNAs to inhibit *lacZ* were constructed and used to demonstrate that no differences arose from targeting our control to a nonsense perturbation (*rfp*) or a gene irrelevant to fitness (β-*galactosidase* encoding *lacZ*) (Supplementary Fig. 7). All subsequent experiments use nonsense *rfp* perturbations as controls.

6.3.2 Increasing the number of gene perturbations reduces competitive fitness

We evaluated the fitness impacts caused by each CHAOS construct by competing these experimental strains with the fluorescent control strain $C_{mcherry}$ during exposure to 0.005 µg/mL ciprofloxacin, a sub-minimal inhibitory concentration (MIC) allowing for moderate growth while still imparting selective pressure. We chose ciprofloxacin as it is a clinically-relevant antibiotic treatment which selects for resistant populations at very low concentrations[281] via specific mutations in the *gyrA* gene[282], thus allowing us to assess adaptation of strains. The CHAOS and controls strains were competed for one day followed by plating on solid media. Fitness impacts were quantified by measuring the relative changes in colony forming units of both control ($C_{mcherry}$ – red colonies), and the competed CHAOS strain before and after exposure to sub-MIC of ciprofloxacin using fluorescence imaging (see methods). No significant differences were observed between either control strains Control or CCCC (exhibiting fitness of 0.99 ± 0.07 and 0.98 ± 0.14 respectively).

Individual gene perturbations of stress response genes either had no statistically significant impact (*soxS, tolC*) or increased fitness (*mutS, recA*) in relation to the control

(Fig. 2c). A striking trend emerged as we multiplexed these four perturbations. While the average fitness of individual perturbations was improved (1.22 ± 0.30, P = 0.003), their benefits were abated when combined in pairs (0.98 ± 0.27, P = 0.88), triplets (0.78 ± 0.23, P = 0.002), and all at once (0.87 ± 0.23, P = 0.10) (Fig. 2c). Only one double perturbation was beneficial (*mutS-recA*), while another two (*soxS-tolC* and *tolC-recA*) resulted in significant fitness losses. Furthermore, half of the triple perturbations reduced fitness (*mutS-tolC-recA* and *soxS-tolC-recA*).

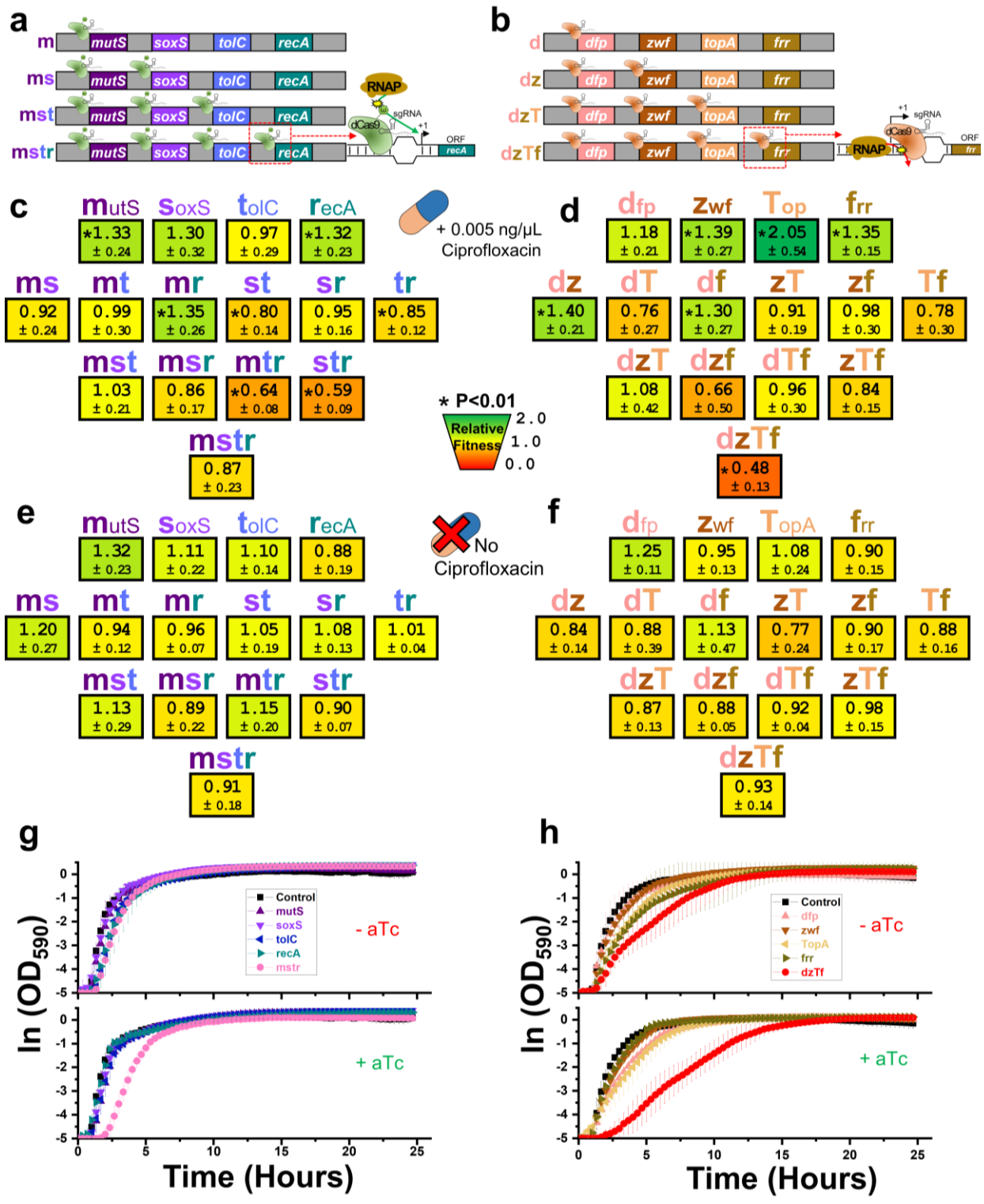


Figure 6.2 Negative fitness from multiple perturbations of gene expression

a Multiplexed CRISPR activation of stress response genes using aTc-inducible dCas9-ω engineered to bind upstream ~80-100 nt upstream of the +1 transcription start site. **b** Multiplexed CRISPR inhibition of conserved genes using aTc-inducible dCas9

engineered to bind overlapping the +1 transcription start site. **c-d** Fitness relative to control strain ($C_{mcherry}$) after combinatorially increasing expression of stress response genes (**c**) or decreasing expression of conserved genes (**d**) during exposure to 0.005 µg/mL ciprofloxacin and 10 ng/mL aTc in LB medium. Relative fitness is listed below each strain name, followed by the standard deviation (n = 8). Asterisks indicate significant fitness differences in relation to strain "C_{ontrol}" grown under the same conditions and exhibiting a competitive fitness of 0.99 ± 0.07 ($P < 0.01$, two-tailed type II *t*-test). No significant differences were observed between strain C_{ontrol} and another control strain "CCCC" harboring four nonsense gene perturbations and exhibiting a competitive fitness of 0.98 ± 0.14. **e-f** Fitness relative to control strain ($C_{mcherry}$) from combinatorially increasing expression of stress response genes (**e**) or decreasing expression of conserved genes (**f**) during exposure to no ciprofloxacin and 10 ng/mL aTc in LB medium. Relative fitness is listed below each strain name, followed by the standard deviation (n = 3 for conserved gene strains, n=4 for stress response gene strains). No strain was significantly different than the control strain "CCCC" with four nonsense gene perturbations grown in the same conditions and exhibiting average fitness of 0.93 ± 0.23 (n=3, P < 0.01, two-tailed type II *t*-test). **g-h** Growth of strains harboring stress response perturbations (**g**) or conserved gene perturbations (**h**) either individually or all at once during exposure to 10 ng/mL aTc in LB medium. Error bars represent standard deviation of five biological replicates.

Perturbation of conserved genes resulted in a similar but even more drastic trend. Three of these perturbations significantly improved fitness (*zwf, topA,* and *frr*), with inhibition of *topA* providing a strong fitness benefit (Fig. 2d). As before, while the average fitness of these four perturbations was significantly improved (1.49 ± 0.44, $P = 8\times10^{-5}$), fitness decreased when these were combined in pairs (1.02 ± 0.34, $P = 0.71$) or three at a time (0.88 ± 0.38, $P = 0.29$). Combining all four perturbations resulted in a particularly severe diminishment of fitness (0.48 ± 0.13, $P = 1\times10^{-11}$), suggesting drastic epistatic effects (Fig. 2d). Collectively, these data support our hypothesis and demonstrate that fitness loss results from combinations of gene perturbations during antibiotic exposure.

We investigated if the inclusion of ciprofloxacin impacted these results by repeating the experiment without ciprofloxacin (Fig. 2e-f). The overall impact of perturbations was diminished under these conditions, and no strain exhibited a competitive fitness

significantly different than the control (CCCC, 0.93 ± 0.23). These strains did exhibit an apparent trend towards lower fitness upon multiplexing; the average fitness of individual stress response gene perturbations (1.10 ± 0.24) decreased upon combining perturbation in pairs (1.04 ± 0.16), triplets (1.02 ± 0.23), and all at once (0.91 ± 0.18) (Fig. 2e). The same was true for multiplexing individual conserved gene perturbations (1.05 ± 0.20) into pairs (0.90 ± 0.27), triplets (0.91 ± 0.10), and all at once (0.93 ± 0.14) (Fig. 2f).

Growth curves of individual and four gene perturbations were also examined in the absence of ciprofloxacin (Fig. 2g-h). No growth changes were observed for any of the stress response gene perturbations in the presence or absence of aTc, while growth of strain mstr did diminish upon induction (Fig. 2g). Similarly, only minor impacts on growth emerged due to perturbation of conserved genes individually, while significant growth defects were present during induction of strain dzTf (Fig. 2h). All strains grew to similar maximum ODs by the end of 24 h. The growth impacts observed in the absence of aTc is likely due to leaky expression from the tet-promoter driving dCas9 and dCas9-ω expression. We quantified a ~100-fold increase in dCas9 expression upon aTc induction, with maximum induction occurring around 3.125 ng/mL (Supplementary Fig. 8), less than the 10 ng/mL used in competition experiments.

Strain growth was analyzed in M9 minimal media to exacerbate potential growth impacts caused by gene perturbations. We varied concentrations from 0 to 50 ng/mL to parse the dCas9 response to induction. At high concentrations of aTc, we observed slight growth rate impacts on control strains, as would be expected for growth in minimal media with two antibiotics for maintaining two plasmids (Supplementary Fig. 9). Again, no significant growth impacts were observed for any of the stress response gene

perturbations (Supplementary Fig. 10). Slight growth defects were observed due to conserved gene perturbations of *topA* at high aTc concentrations (Supplementary Fig. 11). Perturbations of strains mstr and dzTf demonstrated significant growth defects at high aTc concentrations that were maintained up to the end 24 hours of growth (Supplementary Fig. 11). This resulted in significant reductions in growth rates at aTc concentrations used during competition (Supplementary Fig. 12).

We also investigated whether growth in microplate cultures impacted our overall experimental outcomes, as growth in such conditions has frequently been correlated with oxidative stress due to poor oxygenation[283]. We found that batch growth competition resulted in similar conclusions as observed in microplate growth (Supplementary Fig. 13).

Finally, we investigated whether inhibition or activation of gene expression had any impact on the phenomenon observed in Fig. 2c-d. For this, we created another four-perturbation strain activating expression of *mutS* and *soxS* while inhibiting expression of *topA* and *frr.* We tested the competitive fitness of this strain during ciprofloxacin exposure. While three of these perturbations significantly improved fitness, perturbation of all four simultaneously resulted in neutral fitness (1.09 ± 0.10) (Supplementary Fig. 14). This strain also grew slower than the control strain, corroborating the notion that these perturbations interacted detrimentally upon multiplexing.

Taken together, these results demonstrate that individual gene perturbations were not detrimental to the fitness of *E. coli*. Only upon multiplexing these perturbations did significant growth and fitness impacts emerge, which were markedly more pronounced during exposure to ciprofloxacin. Thus, a clear trend towards lower fitness emerged upon multiplexing of gene perturbations.

6.3.3 Strong trends towards negative epistasis emerge from combining gene perturbations

We next quantified epistasis between simultaneous perturbations by calculating deviations between the measured fitness of multiple perturbation strains, and their expected fitness based upon single perturbation (see methods). Comparing these fitness values, no deviation between expected and actual fitness of strain CCCC was observed, while multiple perturbation strains clearly exhibited lower fitness than was expected (Fig. 3a). This trend correlated into significant negative epistasis in half of the double perturbation strains and all but one of triple and quadruple gene perturbation strains (Fig. 3b). Notably, the only gene pairs known to interact (*mutS-recA* and *soxS-tolC*) did not demonstrate significant epistasis, indicating that direct interaction is not required to produce negative epistatic effects. The degree of negative epistasis also appeared to increase as more genes were perturbed. Inhibition of conserved genes resulted in statistically greater levels of negative epistasis than activation of stress response genes for triple perturbation constructs ($P = 0.01$). The high degree of negative epistasis across both sets appears compounded by sign-epistasis, wherein individually beneficial perturbations become deleterious once combined. Raw epistasis values and significance are presented in Supplementary Table 2.

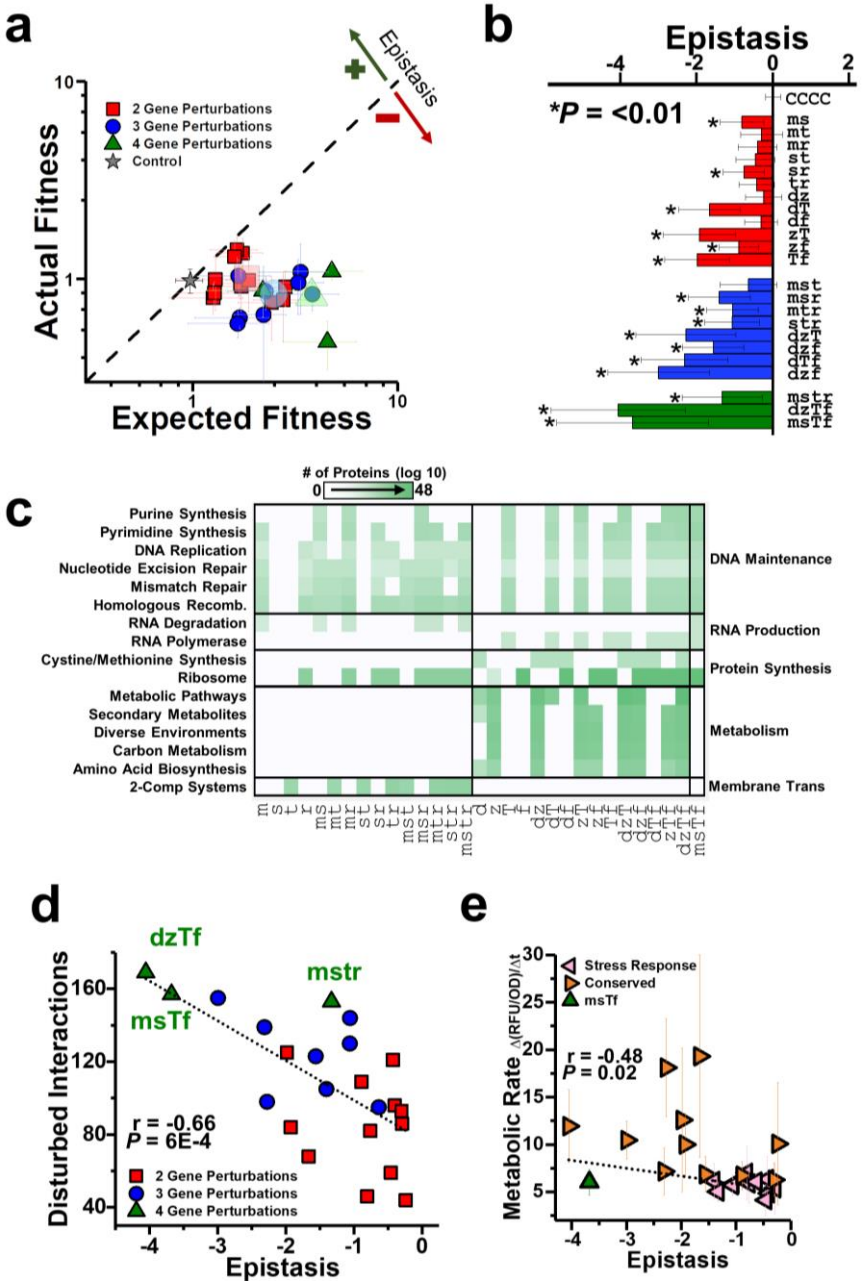


Figure 6.3 Epistasis resulting from two or more gene perturbations

a The relationship between expected and actual relative fitness of each strain harboring multiple gene perturbations. Centroid of each group (based on the number of genes perturbed) is shown by the larger transparent symbol. The dashed diagonal line indicates theoretical results if no epistasis was present. **b** Calculated epistasis of each strain (abbreviated as in Fig. 2). Error bars indicate standard deviation (n = 8). Raw expected fitness and epistasis values are presented in Supplementary Table 2. Asterisks indicate significant negative epistasis in relation to the null hypothesis of zero epistasis ($P < 0.01$, Student's *t*-test). **c** Counts of proteins directly impacted by gene perturbations, separated by functional classes as annotated by the Kyoto Encyclopedia

of Genes and Genomes (KEGG). Only statistically significant ($P < 0.05$) over-represented KEGG pathways are reported. **d** Relationship between epistasis and the cumulative amount of direct protein interactions disturbed by each perturbation set. A linear fit is included, with Pearson correlation coefficient (r) and its significance (*P*). **e** The metabolic activity of all strains after sub-MIC ciprofloxacin exposure in LB medium quantified using fluorescence change from Resazurin dye. Fluorescence was normalized to final ODs (580nm) of the respective replicate. Data represent the average of at least three biological replicates. A linear fit of the metabolic rate's relationship to epistasis was performed using ANOVA associated r and *P* values presented.

We investigated the functional processes influencing epistasis. We started by identifying all proteins known to directly interact with the gene(s) disrupted by CRISPR perturbation(s) as characterized by the STRING database[280]. Functional differences between the affected pathways from each set could explain these differences, as conserved gene perturbations disrupted pathways more central to cell survival. Exploring the functionality of each gene within the impacted networks, we observed that all simultaneous perturbation constructs impacted pathways related to DNA maintenance, RNA production, and protein synthesis, but metabolism and membrane transport were unique to conserved gene perturbations and stress response perturbations respectively (Fig. 3c). As expected, introducing more primary CRISPR perturbations introduced more cascading downstream impacts on protein partners within the *E. coli.* Interestingly, we observed a strong correlation ($P = 6\times10^{-4}$) between the total amount of affected partners and the degree of negative epistasis exhibited by each CHAOS strain (Fig. 3d). This suggests that as more interactions are disturbed from homeostasis, the greater the impact of epistasis on the cell's overall health, likely due to increased disruption of homeostasis.

As central metabolic pathways were impacted only by conserved gene perturbations, we quantified metabolic rates of each strain after 20 hours of exposure to sub-MIC levels of ciprofloxacin (Fig. 3e). We observed greater metabolic activity within

the conserved set of perturbations (P = 8×10^{-7}), and a significant correlation between metabolic activity and the degree of epistasis (r = -0.48, P = 0.02). A potential mechanism for this could be reduced efficacy of resource dedication towards surviving antibiotic exposure. This is supported by previous work correlating increased metabolic rates to potentiation of bactericidal antibiotics[284] and another study correlating antibiotic efficacy to the presence of metabolites[285]. Furthermore, 13 of the affected metabolic genes in strain dzTf are essential. Adapting bacteria exhibit less stochastic expression in essential genes[17], and negative epistasis has been proposed as a mechanism constraining this heterogeneity[217] due to essential genes exhibiting stronger overall genetic interactions[286]. Indeed, genetic interactions are statistically greater in the conserved gene set than stress response set (Supplementary Fig. 15, P = 7×10^{-10}). Collectively, these studies appear to corroborate our finding that accumulating gene expression deviations in diverse cellular pathways has a fundamental tendency to influence fitness detrimentally in an epistatic fashion.

6.3.4 Multiplexed gene perturbations slow bacterial adaptation rates and increase antibiotic susceptibility

To test the hypothesis that induced negative epistasis can restrict the adaptive potential of bacteria, we exposed single and quadruple perturbation strains to ciprofloxacin over three days (D1-3) and quantified changes in MIC (Fig. 4, see methods). The trajectory of MIC change was quantified using Pearson correlation coefficients of linear fits over time (r, see methods). Statistical differences in fits between the control and CHAOS strains were estimated using F-tests (reported as P > F values).

Most single perturbation strains and the controls adapted similarly, with MICs increasing over time. *recA* activation resulted in a significant increase in MIC on D1 ($P = 2\times10^{-5}$) and D3 ($P = 2\times10^{-3}$), and the rate of this increase was significantly faster than the control ($P > F = 9\times10^{-5}$) (Fig. 4a). While *topA* perturbation lowered MIC on D1 ($P = 5\times10^{-15}$), the strain quickly adjusted back to control levels resulting in a statistically faster rate of increase than the control ($P > F = 5\times10^{-7}$) (Fig. 4b). The discrepancy between fitness and initial MIC impacts of individual *topA* perturbation may be due to the aforementioned gene expression dependency on cell phase. As expected, no differences were observed between single and four gene control perturbation strains (Fig. 4c).

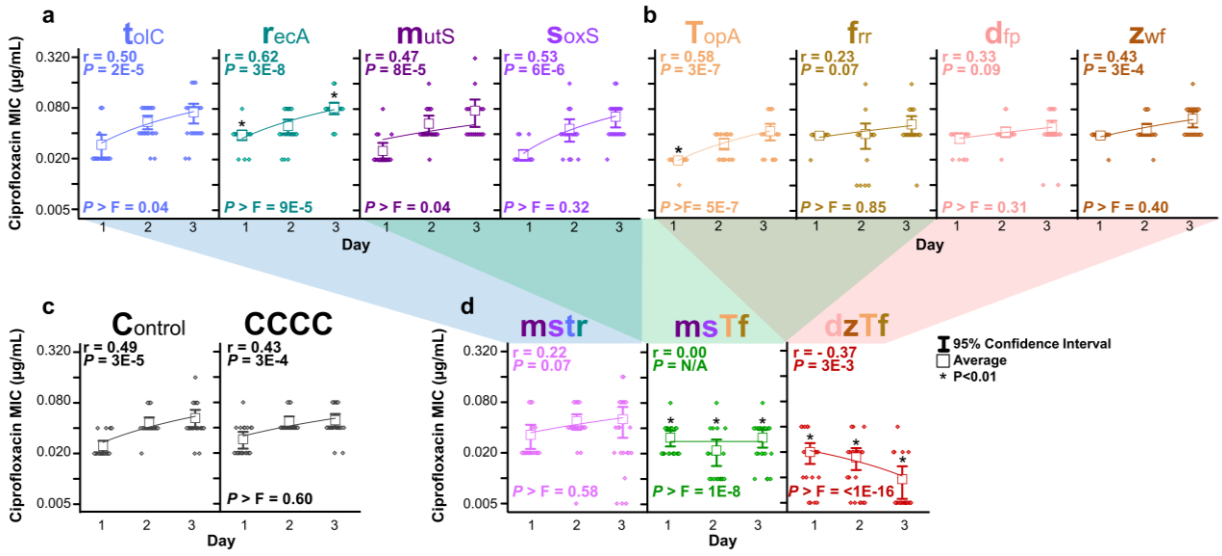


Figure 6.4 Perturbation of multiple genes slows bacterial adaptation

Change in the MIC of ciprofloxacin towards adapting populations at the end of each day of exposure for **a** strains harboring individual stress response gene perturbation constructs for *mutS*, *soxS*, *tolC* and *recA* activation, **b** strains harboring individual conserved gene perturbation constructs for *dfp*, *zwf*, *topA* and *frr* inhibition, **c** control strains, and **d** strains with simultaneous perturbation of four genes corresponding to stress response genes only (mstr), combination of stress response and conserved genes (msTf), and conserved genes only (dzTf). Each box-plot (n = 22, individual data points shown) includes a linear fit with associated Pearson correlation coefficient (r) and its significance (*P*). F-tests were performed for strain's linear fit against the linear fit of strain C_{ontrol} during the same experiment, and the resulting significance is reported as *P* > F. Raw MIC values are presented in Tables S3-5. Asterisks indicate significantly different average MICs in relation to strain C_{ontrol} during the same day and experimental run ($P < 0.01$, two-tailed type II *t*-test).

Experimental strains of quadruple perturbations exhibited striking differences in adaptive trends. Strain mstr (*mutS*, *soxS*, *tolC,* and *recA*) exhibited a positive correlation that was weaker than each individual perturbation (Fig. 4d). Strain msTf (*mutS*, *soxS*, *topA,* and *frr*) exhibited a completely flat MIC trajectory, and always survived statistically lower ciprofloxacin concentrations than the control. Strain dzTf (*dfp*, *zwf*, *topA,* and *frr*) presented the most striking results, as not only were MICs statistically lower than the control, average MIC actually decreased over time (r = -0.37, $P = 3\times10^{-3}$). This was

primarily due to death of fifteen replicates by the end of adaptation. Nevertheless, excluding replicates that died during the course of the experiment showed a neutral trajectory of MIC over time (r = -0.09, *P* = 0.56), indicating that the living population was adapting slower (Supplementary Figs. 16-17). The failure of strain dzTf to adapt to ciprofloxacin exposure was reproducible across multiple experimental runs (r = -0.42, *P* > F = 7×10^{-15}) (Supplementary Fig. 18). Strains exhibiting greater negative epistasis (dzTf and msTf) also adapted at slower rates, suggesting a correlation between the degree of epistasis and the rate of adaptation. Raw MIC values for all strains discussed above are presented in Supplementary Tables 3-5. These results demonstrate the potential of employing CHAOS to introduce epistasis into an organism and constrain rates of bacterial adaptation to antibiotics over time.

To determine if differences in MICs translated to changes at the genetic level, we sequenced each strain at the end of three days of ciprofloxacin exposure. We focused specifically on *gyrA*, as mutations in this gene (S83L and D87Y) have been found as the first genetic changes during the evolution of ciprofloxacin resistance[282]. The vast majority of strains exhibited no *gyrA* mutations (Supplementary Table 6). In total, only four isolates exhibited any *gyrA* mutations. These included one replicate of the control, and one replicate each of strains with individual perturbation of *mutS*, *soxS,* and *topA*. Increased perturbations did not appear to bias cells to mutate more, as no replicates of mstr, msTf, and dzTf exhibited mutations in *gyrA*. Mutation fluctuation assays corroborate this, revealing similar mutation rates across all strains with the exception of individual *mutS* and *soxS* perturbation (Supplementary Fig. 19). Due to similar mutation rates, differences

in adaptive trajectories across strains appear to be driven primarily by changes in gene expression.

We further explored strain adaptability by looking at the sustainability of MIC changes. We again adapted each of the individual conserved gene perturbations, as well as strains dzTf and the control, for three days (Fig. 5a). We found similar results as before, with strain dzTf adapting particularly poorly. After three days of exposure, all strains were removed from ciprofloxacin exposure for two days to reset the phenotypic state, after which they were re-exposed to ciprofloxacin gradients. Despite growth in the absence of ciprofloxacin for two days, every strain survived to the same levels of ciprofloxacin it had on day three. This suggests that the bacteria exhibited sustainable adaptive changes in response to ciprofloxacin, except for the bacteria exposed to multiplexed perturbations inducing negative epistasis.

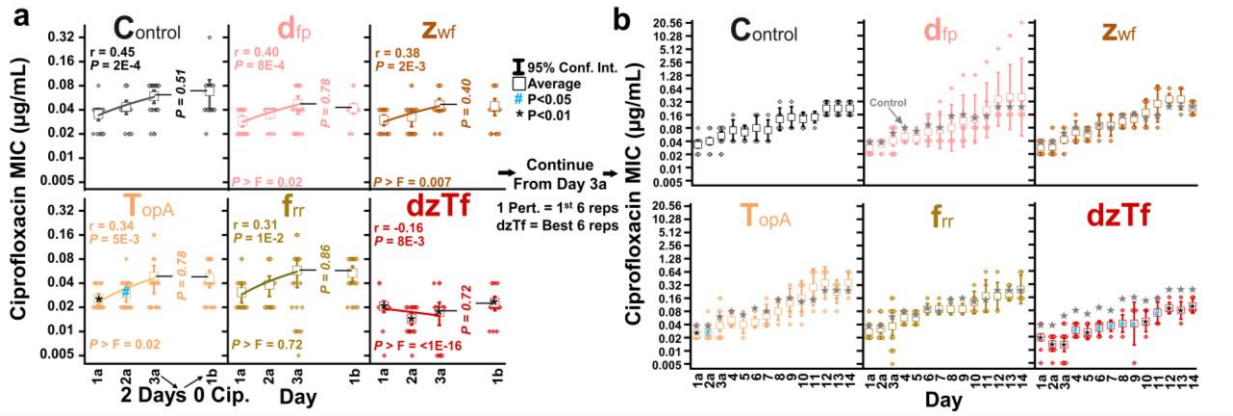


Figure 6.5 The impact of multiplexed conserved gene perturbation is maintained in the absence of ciprofloxacin and for longer periods of adaptation

a Change in the MIC of ciprofloxacin towards adapting populations of strains harboring the noted perturbation constructs Each box-plot (n = 22, individual data points shown) includes a three-day linear fit with associated Pearson correlation coefficient (r) and its significance (*P*). F-tests were performed for strain's linear fit against the linear fit of strain C_{ontrol} during the same experiment, and the resulting significance is reported as *P* > F. At the end of the third day (3a), samples were inoculated into LB with induction and grown for two days in the absence of ciprofloxacin. The MICs of these samples were measured again after one day (1b) of ciprofloxacin exposure. **b** The experiment was continued for six replicates of each strain at the end of three days of exposure (3a) for day four to day 14. The first six replicates for all strains besides dzTf, which continued the best six replicates reaching the highest MICs. Stars are overlain on each graph to show the average MIC of the control strain. All asterisks (*) and hashtags (#) indicate statistical difference from the control strain on the same day of the experiment (*P* < 0.01, two-tailed type II *t*-test).

Finally, we also examined the adaptive potential of these strains in a longer, clinical relevant timeframe of two weeks (Fig. 5b). Six replicates of each of the six strains from the above experiment were taken at the end of day three, and continued for another 11 days of adaptation to ciprofloxacin. We chose the first six replicates of each of the individually perturbed strains to continue with, and biased our experiment by picking the six best replicates of strain dzTf to continue with adaptation. This was done to see if a subset of the dzTf population was able to escape the epistatic effect and evolve resistance to ciprofloxacin. Despite this bias, strain dzTf never managed to reach similar levels as

the control strain throughout the entire experiment. The population did begin surviving higher ciprofloxacin levels around day eight of the experiment, suggesting that adaptation was eventually possible. But the speed at which this strain evolved resistance was clearly diminished, suggesting that CHAOS restriction of adaptation is applicable in clinically relevant timeframes.

6.3.5 Multiplexed peptide nucleic acid gene perturbations increase chloramphenicol susceptibility of a clinical multi-drug resistant isolate of E. coli

To demonstrate the therapeutic potential of CHAOS we multiplexed gene perturbations against a clinical isolate of carbapenem-resistant Enterobacteriaceae (CRE) *E. coli*, a bacterial pathogen recently designated as priority 1 critical class by the World Health Organization[287]. Characterization of this isolate using the 2016-2017 Clinical & Laboratory Standards Institute (CLSI) sensitive/resistant breakpoint values[149] showed resistance to at least eight classes of antibiotics[288,289]. For this study, we tested the response of this strain to chloramphenicol as it exhibited greater than 8-fold higher MIC (>256 µg/mL) than the corresponding CLSI breakpoint of 32 µg/mL (Fig. 6a). We targeted four new genes, allowing us to confirm if CHAOS is generalizable. Based on the success of CHAOS strain dzTf, we chose to target essential genes representing diverse cellular pathways. These genes included *folC*, an H2-folate synthetase involved in folate biosynthesis[290], *ffh* encoding a signal recognition particle protein gene essential for protein translocation[291], *gyrB* encoding gyrase subunit B important for transcription, and an essential noncoding small RNA *fnrS.*

As delivery of CRISPR systems into bacteria is still difficult to achieve therapeutically, we employed an alternative gene expression repression strategy based on 12-mer peptide nucleic acids (PNAs, abbreviated as α-*gene*). These PNAs target and bind to the translation start codon of these genes, thus inhibiting translation of these genes' mRNAs into protein[95] (see methods). We chose concentrations of PNA at 2.5 μM at which individual perturbations had minimal to no effect on cell growth of standard MG1655 *E. coli*. These PNAs were conjugated via an O-linker to a cell-penetrating peptide (CPP) motif for direct intracellular delivery. We exposed the clinically isolated CRE *E. coli* to these PNAs individually and in combination for 24 hours, before plating in both the presence and absence of chloramphenicol (Fig. 6b).

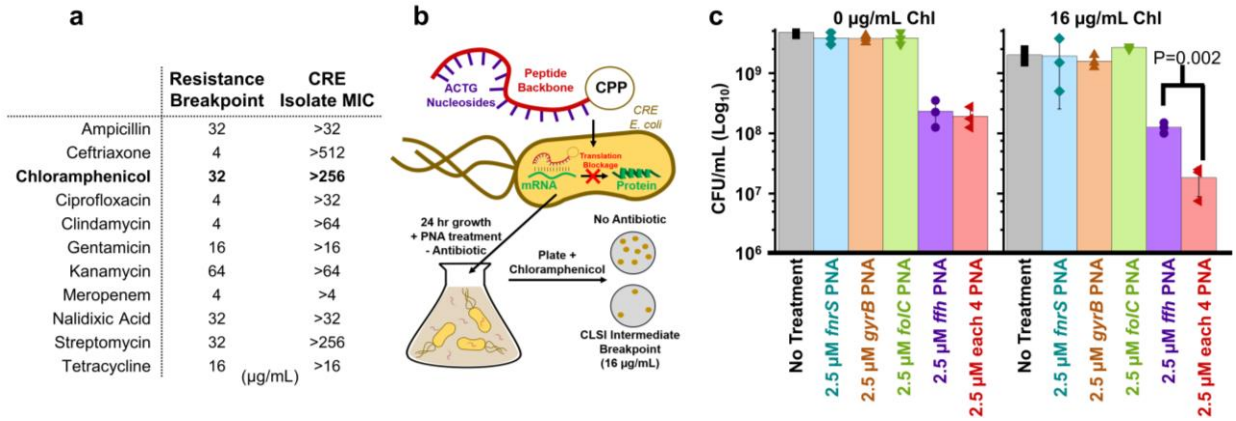


Figure 6.6 CHAOS increases the antibiotic susceptibility of clinically isolated CRE *E. coli*

a A CRE isolate of *E. coli* exhibiting resistance to at least 11 antibiotics above CLSI breakpoint levels was isolated from a clinical infection. We focused on applying CHAOS induced epistasis to re-sensitize this isolate to chloramphenicol. **b** A new set of four universally conserved bacterial genes were perturbed using PNA to demonstrate applicability outside of CRISPR interference and towards clinically relevant infections. PNA structure consists of a peptide backbone connecting nucleosides analogous to DNA and linked to a cell penetrating peptide. These molecules are able to enter bacteria and anneal tightly to analogous mRNA sequences, allowing for targeted blockage of protein translation. Chloramphenicol-resistant CRE *E. coli* was exposed to 2.5 μM of four unique PNAs either individually or in combination (for a total concentration of 10 μM PNA) for 24 hours, after which cells were plated on both plain LB agar, as well as clinically-relevant levels of chloramphenicol to determine viable cells. **c** CFU analysis of CRE *E. coli* after exposure to PNA treatment demonstrates CHAOS's effectiveness. Exposure to PNA *ffh* resulted in a ~16-fold reduction in viable cells with respect to no PNA treatment, while the remaining PNAs exhibited largely no effect under both conditions. Combination of all 4 PNAs exacerbated chloramphenicol's toxicity and gave rise to ~110-fold reduction in viable cells with respect to no PNA treatment in an apparently epistatic fashion even at sub-resistance levels. P values were calculated using two-tailed type II t-test.

We observed that in absence of chloramphenicol, the total amount of viable cells of the isolate was affected only by α-*ffh* (16-fold decrease with respect to no PNA treatment), with no apparent exacerbation of α-*ffh*'s toxic effect upon multiplexing with other PNAs (Fig. 6c). When these same cultures were plated in the presence of 16 μg/mL of chloramphenicol, an intermediate CLSI-breakpoint level, similar results were observed for all of the individual PNA treatments. However, multiplexing of all four PNAs

significantly reduced the viability of the clinical isolate by nearly 110-fold with respect to no PNA treatment and resulted in a significant reduction (P=0.002) compared to the best individual perturbation. We note that overall PNA concentration was increased under this condition, leading to the potential of slightly increased antibiotic permeability (Supplementary Fig. 20). However, such impacts were relatively small, and disappeared after 15 hours of growth. Taken together, this data provides evidence that CHAOS can be used to increase the susceptibility of highly drug-resistant clinical isolates in an apparently epistatic fashion.

6.4 Discussion

This work provides the framework for CHAOS, a new strategy to slow bacteria's ability to adapt to antibiotics by employing multiplexed perturbations of gene expression to induce negative epistasis. The influence of gene expression epistasis on evolutionary trends is a relatively recent and virtually unexplored idea[217,218,292]. The data presented here suggest that excessive divergence from basal expression in multiple genes induces negative epistasis even when no direct genetic interactions exist. This effect appears to be influenced by the number of perturbations, as well as the type of pathways affected (Fig. 3).

An intuitive explanation for this phenomenon is that increasing perturbations generally leads to a breakdown in homeostasis, as shifts in independent pathways create cascading detrimental impacts. For example, despite no direct interaction between *topA* and *zwf*, both plausibly influence DNA folding rates through different mechanisms – *topA* through altered topoisomerase expression, and *zwf* through altered NADPH production,

influencing flux through the thioredoxin pathway that regulates *gyrA* expression[293]. Such collateral effects help explain why independent perturbation of *topA* and *zwf* appears to increase fitness during ciprofloxacin exposure, but their combined perturbation results in negative epistasis. Furthermore, this disruption of homeostasis via gene expression changes could explain why essential genes are conspicuously restricted in heterogeneity[217], particularly after stress exposure[17].

Our previous results also corroborate this notion, wherein a smaller set of five strains harboring multiplexed inhibitions and activations of two to three stress response genes tended to exhibit negative epistasis when exposed to various stressors including both antibiotics and biofuels[19]. This raised the idea that gene expression based epistasis might be a fundamental phenomenon, paving the way for this work in which we significantly expand this idea.

Artificially engineering bacterial gene expression to increase antibiotic susceptibility is a relatively new idea. Previous work has shown multiplexing transcription factor overexpression can successfully re-sensitize antibiotic resistant bacteria[294], but is limited to mass alterations of genetically related genes regulated by the same native processes bacteria have naturally evolved. In contrast, CHAOS can theoretically be applied to modify any set of genes and can be interfaced with either a CRISPR or PNA platform.

One of the major limitations to implementing CHAOS in a clinical setting is the delivery of these gene perturbation systems into the target cells, especially for intracellular infections. Significant work is being undertaken to improve the therapeutic delivery of CRISPR systems[65,80,81,85,295,296] and PNAs[97,105,297], and there is reasonable potential that

these approaches could be applied to making CHAOS a viable therapy. This is best demonstrated by our application of CHAOS to increase the susceptibility of a clinically isolated multi-drug resistant CRE *E. coli.* Despite high levels of resistance, we were able to increase its susceptibility to chloramphenicol with multiplexed perturbations of four essential genes.

Hindering the evolution of antibiotic resistance is a major goal of antibiotic synergy in combination therapy approaches[298], and engineered negative epistasis affords an ancillary but functionally separate strategy to achieve this goal. Exploiting genetic interactions is ripe for use in therapeutic strategies, as evidenced by the growing interest in synthetically lethal mutations that are detrimental to cancer mutated cells but benign to healthy cells[299]. Future studies involving a systematic investigation of transcriptome-wide epistasis would help in this regard by identifying optimal gene combinations to achieve detrimental interactions with minimal invasiveness. Such studies will help to confirm whether the sign of perturbation (activation versus inhibition) or the degree of perturbation could influence the level of emergent epistasis. Overall, our results demonstrate that applying a more sophisticated understanding of gene expression enables one to induce negative epistasis to gain control over adaptation rates, and could preserve the efficacy of last-resort antibiotics when employed in co-therapy strategies.

6.5 Materials

6.5.1 Plasmid and strain construction

A list of plasmids and primers used in this study can be found in Supplementary Tables 1 and 7 respectively. A two-plasmid system was utilized to induce native gene expression

perturbation; the first plasmid encoded a sgRNA target sequence, while the second encoded either dCas9 or dCas9-ω for gene inhibition and activation respectively. Addgene plasmid 44249 was used directly for targeting *rfp* inhibition (the “Control” – C) and served as the starting plasmid for creating all subsequent sgRNA plasmids. Addgene plasmid 44251 was used directly for providing dCas9, while the previously constructed pPO-dCas9ω[19] plasmid was used directly for providing dCas9-ω. New sgRNA target plasmids were created by replacing the RFP-targeting sequence in 44251 with new gene sequences specific to the target of interest. This was accomplished by designing unique forward primers flanked with a *SpeI* restriction site and encoding the new target sequence. A common reverse primer flanked with *ApaI* was used alongside these primers to perform PCR amplification with Phusion High-Fidelity DNA Polymerase (New England Biolabs) of DNA inserts, which were subsequently digested with Cutsmart SpeI and ApaI (New England Biolabs) alongside 44251 backbone. Ligations of these pieces were performed using T4 DNA Ligase (Thermo Scientific), which were subsequently transformed into electrocompetent NEB 10-β. Transformants were minipreped using Zyppy Plasmid Miniprep Kit (Zymo Research Corporation) and submitted for sequencing to confirm successful insertion (GENEWIZ). sgRNA plasmids targeting individual genes were used to construct sgRNA plasmids targeting two or more genes via Gibson Assembly, for which a common forward and reverse primer was used to amplify the first sgRNA target plasmid while introducing overhangs downstream of the terminator sequence following the first target. A common set of primers were then utilized to amplify sgRNA targets to the second, third, and fourth targets depending on the intended number of final sgRNA targets. A batch Gibson reaction was performed at 50 °C for 3 h with T5

exonuclease (New England Biolabs), Phusion polymerase and Taq ligase (New England Biolabs) on this one backbone and one to three inserts to stitch all pieces together (Supplementary Fig. 4). sgRNA-C-mCherry was constructed by amplifying constitutively expressed mCherry from pFPV-mCherry (Addgene 20956) and inserting into 44251 upstream of the *rfp* sgRNA sequence. Final experimental sgRNA plasmids were transformed into chemically competent *E. coli* strain K-12 MG1655 (ATCC 700926) harboring either 44249 or pPO-dCas9ω if the target was meant to inhibit or activate expression respectively. This process was used to construct all control and experimental strains used in the study.\

6.5.2 Media and culture conditions

All cultures were grown in Lennox Luria-Bertani Broth (LB) (Sigma-Aldrich), with the exception of RT-qPCR samples and certain samples for growth calculations which were grown in M9 minimal media (5X M9 minimal media salts solution from MP Biomedicals, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$ supplemented with 0.4 % weight/vol glucose). Plates and media were supplemented with ampicillin (100 µg/mL) or chloramphenicol (35 µg/mL) to maintain selection of sgRNA plasmids or dCas9/dCas9-ω plasmids respectively. aTc was used to induce CRISPR expression at a final concentration of 10 ng/mL, except where otherwise noted. The authors also note that the aTc-inducible promoter driving expression of dCas9 is not $P_L$tetO-1 as originally reported[248], but rather a tet-promoter variant with only one Tet binding site highly similar to the original tet-promoter, indicating that slightly higher leaky expression is expected of dCas9 and dCas9-ω. All cultures were grown at 37 °C, with shaking at 225 rpm. Cultures for competition were grown in 200 µL

cultures in 96 well conical-bottom microplates. Cultures for RT-qPCR and for batch growth comparisons to microplates were grown in 3 mL cultures. Cultures for CFU and MIC screens were grown in 100 µL cultures in 384 well flat-bottom microplates. Cultures for growth rate calculations were grown in 100 µL cultures in 384 well flat-bottom microplates in a GENios plate reader (Tecan Group Ltd.) operating under Magellan software (version 7.2) with 16.6 min of shaking before measurement of optical densities at 590 nm absorbance every 20 min.

6.5.3 Quantitative reverse transcription PCR

The degree of gene expression perturbation was confirmed by subjecting biological triplicates of each individual gene perturbation to RT-qPCR, as well as constructs perturbing four genes simultaneously and the control strain targeting *rfp*. Cultures were inoculated from individual colonies and grown for 20 hours overnight in 3 mL M9 cultures and subsequently diluted 1:100 the following morning into 3 mL of fresh media containing aTc. These cultures were grown for 8 hours before RNA extraction using the GeneJET RNA Purification Kit (Thermo Scientific) and purification using the Turbo DNA-free kit (Ambion). Separate biological replicates of *topA* perturbation and the single target *rfp* control strain were also grown for 24 hours to stationary phase and collected as before. Purified RNA was used to create cDNA using the Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific). Technical duplicates of each replicate were subjected to RT-qPCR reactions from the Maxima SYBR Green qPCR Master Mix (Thermo Scientific) using 2 ng of cDNA in 12.5 µL reactions run on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems) in the CU Core Sequencing Facility.

Reactions were allowed to run for 40 cycles with Rox normalization. Gene expression changes were calculated using $2^{-\Delta\Delta Cq}$ values calculated from averages of technical duplicates, relative to the control strain targeting *rfp*.

6.5.4 Competition fitness assay

The fitness of each perturbed strain was calculated by competing said strain against the red fluorescent control strain "$C_{mCherry}$". A total of eight biological replicates were inoculated from colonies into 200 µL in 96 well plates and grown overnight for 16 h with selection. Cultures were then diluted 1:100 into 200 µL of media with selection and 10 ng/mL aTc to induce gene perturbation and grown for another 24 hours. Competition was initiated by diluting cultures 1:100 and mixing equal cell ratios of the red control strain with each experimental strain into 200 µL of media containing selection, 10 ng/mL aTc and 0.005 µg/mL ciprofloxacin. To determine starting ratios of each strain, two µL were diluted into 198 µL of water, from which 10-fold serial dilutions up to 1:100,000 were created each using a total volume of 200 µL for each dilution. 50 µL of both the 1:10,000 and 1:100,000 dilutions were plated to determine CFUs (and by extension starting ratios) of each strain. The remaining 198 µL of culture not used for serial dilutions was grown for another 24 h, diluted 1:100 into the 198 µL of the same media, and grown again for 24 h. At the end of this growth period, CFUs for each strain were determined using the previously outlined dilution scheme. Plates were incubated for 24 hours after inoculation. Two images were taken of each plate with fluorescence activation at 540 nm, one with emission filtering at 590 nm and the other without. These images were overlaid to facilitate

colony counting. Colony counts were used to determine fitness values (ω) using the standard Malthusian fitness equation[227]:

$$\omega = \ln(N_{E1}\times 100^2/N_{E0})/\ln(N_{C1}\times 100^2/N_{C0})$$

where the variables are defined as follows: “N” – CFU, “E” – experimental strain, “C” – control strain, “1” – after exposure, and “0” – before exposure. Fitness values were calculated as such for all experimental strains, as well as the control strains targeting *rfp* with one or four sgRNAs and expressing no fluorescence. This same protocol was repeated in the absence of ciprofloxacin for all combinations of conserved gene perturbations

6.5.5 Epistasis calculations

Expected fitness values (ω_e) for strains with perturbation of two or more genes were calculated assuming a multiplicative model as follows:

$$\omega_e = \prod_{i=1}^{n} \omega_i$$

where n expands to all sets of genes perturbed. For instance, ω_e of strain dzf would be calculated as the product of fitness from each individual gene perturbation ($\omega_d\times\omega_z\times\omega_f$). Epistasis (E) was calculated as the difference between measured fitness and expected fitness ($E=\omega-\omega_e$).

6.5.6 Growth assay

To demonstrate growth phenotypes, biological triplicates of each strain were inoculated from individual colonies into 150 µL of LB containing selection in a conical 96 well

microplate and grown for seven hours. After initial growth, one µL of each culture was used to inoculate two mL of M9 or LB media containing selection and grown for 16 h. The following morning, each culture was diluted 1:100 into 100 µL M9 or LB media cultures containing selection and a variable concentration of aTc from 0 to 50 ng/mL. These cultures were grown in 96 or 384 well microplates in a Tecan GENios reader for 24 hours, measuring Optical Density at 590 nm every 20 min.

6.5.7 Resazurin metabolic rate assay

Four individual colonies were cultured overnight (16hrs) in 100 µL of LB supplemented 35 µg/mL of chloramphenicol, and 100 µg/mL ampicillin a 384 well microplate. These cultures were then diluted 1:100 into LB supplemented with cm, amp, 0.005 µg/mL ciprofloxacin, and 10 ng/mL aTc. The microplate was then incubated in a Tecan GENios microplate reader with continuous shaking at 37 °C while measuring optical densities (580nm). After 20 hours of growth, 10x Resazurin (a dye which fluoresces brightly upon interaction with intracellular NADH, thereby quantifying metabolic activity) was added to each well (final concentration of 0.01 mg/mL), and fluorescence (excitation 485 nm and emission 610 nm) was measured in 5-minute intervals. The slope of the curve from the first 5 data points was used to quantify metabolic activity during the initial stages of ciprofloxacin exposure.

6.5.8 Minimum inhibitory concentration assays

MIC assays were performed using 22 biological replicates per strain. Individual colonies were inoculated into 100 µL LB cultures containing selection and grown for 16 h overnight.

The following morning, cultures were diluted 1:50 into 100 µL of fresh media containing 10 ng/mL aTc in 384 well plates and grown another 24 hours. The following day, each replicate was diluted 1:50 into fresh media containing selection, aTc, and a range of ciprofloxacin concentrations including 0, 0.005, 0.01, 0.02, 0.04, 0.08, and 0.16 µg/mL ciprofloxacin to begin the MIC screen. The new 384 well plate containing variable ciprofloxacin concentrations was grown for 24 h (Day 0 to Day 1), after which absorbance was measured at 590 nm. Cultures expressing ODs greater than 0.20 were determined to have survived. The highest concentration at which each replicate survived was used to inoculate the same plate setup as defined previously (Day 1 to Day 2), while the next highest concentration was determined to be the MIC. This process was repeated for one more day to obtain MICs for 22 cultures at the end of each day of growth for all three days. Three replicates of each strain growing at the highest MICs were saved as glycerol stocks for subsequent sequencing.

Samples for re-exposure to ciprofloxacin were diluted 1:50 into fresh media containing 10 ng/mL aTc, chloramphenicol, ampicillin and no ciprofloxacin. These cultures were grown for 24 hours, diluted into fresh media, and grown for another 24 hours. After two days of growth in the absence of ciprofloxacin, these cultures were transferred back into the ciprofloxacin concentration gradient using the aforementioned protocol.

The same batch of 22 cultures used for re-exposure to ciprofloxacin experiments was used for the fourteen-day adaptation experiment. For this, the first six replicates by order of the control strain, as well as individual perturbations of *dfp*, *zwf*, *topA*, and *frr* were continued on for up to 14 days using the same protocol as before, but increasing

the maximum concentration of the gradient up to 20.48 µg/mL ciprofloxacin. For strain dzTf, we biased our experiment by picking the six best replicates as quantified by maximum MIC reached.

6.5.9 Mutation sequencing assay

Glycerol stocks of strains saved after three days of ciprofloxacin exposure were streaked onto LB agar plates with selection and grown overnight. Two colonies from each plate were used to perform colony PCR amplification of *gyrA* in the 1203 bp region surrounding S83 and D87, the most likely regions for mutations conferring ciprofloxacin resistance to arise. PCR samples were purified and submitted for sequencing (GENEWIZ), for a total of six samples per strain.

6.5.10 Mutation fluctuation assay

Mutation rates were estimated using the rifampicin exposure approach outlined by Luria and Delbruck[175]. Individual colonies were grown in three mL LB without selection for 16 hours and subsequently adjusted to normalized ODs with the addition of LB to denser cultures. Each culture was used to dilute 1:10,000 into 33 parallel 100 µL cultures of LB supplemented with 10 ng/mL aTc and grown for 24 hours. Colony forming units were estimated from three replicates on plain LB agar plates, while the remaining 30 cultures were plated on LB agar containing 100 µg/mL rifampicin. Colonies were counted after 48 hours of exposure, and the FALCOR web tool was used to estimate mutation rates[177].

6.5.11 Statistical analyses

Unless otherwise indicated, all *P* values reported were calculated using a standard two-tailed type II student's *t*-test. Pearson correlation coefficients and their corresponding *P* values were calculated using linear fits with no weighting (OriginPro 9.3.226 software). Comparisons of linear fits (*P* > F values) were performed using F-tests of results from linear fits in OriginPro. Grubb's test for outliers was used to remove individual data points prior to calculations where indicated. Standard deviations of expected fitness (σ_e) of simultaneous gene perturbations were calculated from standard deviations of single gene perturbations (σ_i) using error propagation following the equation: $\sigma_e = \omega_e \times \sqrt{\sum_{i=1}^{n} \left(\frac{\sigma_i}{\omega_i}\right)^2}$. Standard deviations of epistasis (σ_E) were also estimated using error propagation following the equation: $\sigma_E = \sqrt{\sigma^2 + {\sigma_e}^2}$ where σ is the standard deviation of measured fitness of the combined gene perturbation strain. We determined whether epistasis values deviated from the null hypothesis (zero epistasis) by performing a one-sample *t*-test to obtain *P*-values (assuming a two-tailed distribution) using the standard formula $t = (E - \mu)/\left(\sigma_E/\sqrt{n}\right)$ where the variables are defined as follows: t is the test statistic, E is the average sample epistasis, μ is the null hypothesis (in this case, zero), σ_E is the estimated standard deviation of Epistasis , and n is the number of replicates (in this case, eight).

6.5.12 Peptide nucleic acid (PNA) design and synthesis

PNA polymers, peptide backbones connecting standard ATCG nucleosides which bind strongly to RNA and act as effective translation inhibitors, were purchased pre-synthesized (PNA Bio Inc.). These polymers were bought conjugated to a cell-penetrating-peptide motif $(KFF)_3K$ connected to the C-terminus by an O-linker and

supplied as a lyophilized powder. Upon arrival, PNA was resuspended in 5% DMSO to a final concentration of 100 µM. PNA sequences were designed to bind to 12 nts surrounding the translation start site of the targeted gene, keeping minimal off-target homology. These PNA sequences from N to C terminus are as follows: $(KFF)_3$K-O-ATACCATGATTAT (*folC*), $(KFF)_3$K-O-CTCTTGCAGGTG (*fnrS*), $(KFF)_3$K-O-GACAATGTTTGA (*ffh*), and $(KFF)_3$K-O-GTTGATGTCGAA (*gyrB*). The nonsense PNA targeting a non-existent RNA sequence was designed using the sequence GAATAAGGGCGA and also included the $(KFF)_3$K motif on the N terminus. This PNA was graciously donated to us from Dr. Teruna Siahaan of the University of Kansas.

6.5.13 Multiplexed PNA treatment of CRE E. coli

Carbapenem-resistant Enterobacteriaceae (CRE) *E. coli* was obtained resistance levels was obtained from the lab of Dr. Nancy Madinger at the University of Colorado Anschutz Medical Campus, and its resistance profile was characterized in previous work[288,289]. For testing CHAOS against this strain, the clinical isolate was plated on cation-adjusted Muller Hinton Broth (caMHB). Three individual colonies were picked from this plate and grown in 3 mL caMHB overnight for 16 hours. Each culture was then diluted 1:100 into a new 50 µL cultures of fresh caMHB in a 384 well plate and grown for another 24 hours. Each culture was then diluted 1:100 again into six new 50 uL cultures of fresh caMHB in a 384 well plate, this time supplemented with either no PNA, 2.5 µM of each PNA individually, or 2.5 µM of each PNA mixed together (for a final concentration of 10 µM PNA). These cultures were then grown for another 24 hours before performing CFU analysis on plates supplemented with either no chloramphenicol or supplemented with 16

μg/mL chloramphenicol. Plates were incubated for 24 hours after plating before performing CFU analysis.

A control experiment testing the effect of increased PNA concentration was also run using the control nonsense targeting PNA. Individual colonies of the CRE *E. coli* were picked and used to inoculate three 3 mL caMHB overnight for 16 hours. Cultures were then diluted 1:100 into fresh 50 μL caMHB supplemented with either 0, 2.5, or 10 μM of the nonsense PNA in both the presence and absence of 16 μg/mL chloramphenicol in a 384 well microplate. These cultures were then grown in a microplate for 24 hours with continuous shaking while tracking ODs.

6.6 Acknowledgments

We thank A. Palmer and J. Barrick for suggestions and advice preparing the manuscript. We thank N. Madinger for providing the clinically isolated CRE *E. coli*. We thank C. Courtney for characterization of the isolate's drug resistance, as well as design of the 4 PNA molecules used. We thank T. Siahaan for providing the nonsense-targeting PNA.

6.7 Author Contributions

P.B.O. and A.C. conceived the study, designed experiments, and wrote the paper. P.B.O., W.T.C., V.B., and M.J.S. constructed plasmids and strains. P.B.O. performed all experiments relating to competition, epistasis, and rate of adaptation. P.B.O and W.T.C. performed mutation fluctuation assays. P.B.O. and V.B. performed qPCR experiments. P.B.O. analyzed all data.

6.8 Supplementary Info

6.8.1 Supplementary Tables

Supplementary Table 6.1 Strains used in this study.

The host strain is MG1655 for all of the following. Guide RNA targets are followed by an "i" or "a" for inhibition or activation respectively.

Strain	Cas9 Phenotype	Guide RNA
Control	dCas9-ω	*rfp*i
CCCC	dCas9-ω	*rfp*i-*rfp*i-*rfp*i-*rfp*i
C-mCherry	dCas9-ω	*rfp*i + mCherry
mutS	dCas9-ω	*mutS*a
soxS	dCas9-ω	*soxS*a
tolC	dCas9-ω	*tolC*a
recA	dCas9-ω	*recA*a
ms	dCas9-ω	*mutS*a-*soxS*a
mt	dCas9-ω	*mutS*a-*tolC*a
mr	dCas9-ω	*mutS*a-*recA*a
st	dCas9-ω	*soxS*a-*tolC*a
sr	dCas9-ω	*soxS*a-*recA*a
tr	dCas9-ω	*tolC*a-*recA*a
mst	dCas9-ω	*mutS*a-*soxS*a-*tolC*a
msr	dCas9-ω	*mutS*a-*soxS*a-*recA*a
mtr	dCas9-ω	*mutS*a-*tolC*a-*recA*a
str	dCas9-ω	*soxS*a-*tolC*a-*recA*a
mstr	dCas9-ω	*mutS*a-*soxS*a-*tolC*a-*recA*a
dfp	dCas9	*dfp*i
zwf	dCas9	*zwf*i
TopA	dCas9	*topA*i
frr	dCas9	*frr*i
dz	dCas9	*dfp*i-*zwf*i
dT	dCas9	*dfp*i-*topA*i
df	dCas9	*dfp*i-*frr*i
zT	dCas9	*zwf*i-*topA*i
zf	dCas9	*zwf*i-*frr*i
Tf	dCas9	*topA*i-*frr*i
dzT	dCas9	*dfp*i-*zwf*i-*topA*i
dzf	dCas9	*dfp*i-*zwf*i-*frr*i
dTf	dCas9	*dfp*i-*topA*i-*frr*i
zTf	dCas9	*zwf*i-*topA*i-*frr*i
dzTf	dCas9	*dfp*i-*zwf*i-*topA*i-*frr*i
msTf	dCas9-ω	*mutS*a-*soxS*a-*topA*i-*frr*i

Supplementary Table 6.2 Raw epistasis calculations

Strain	Expected Fitness		Epistasis		
	Avg	StDev	Avg	StDev	P
CCCC	0.97	0.15	+0.01	0.20	0.912
ms	1.72	0.52	-0.81	0.57	0.005
mt	1.29	0.46	-0.29	0.55	0.174
mr	1.75	0.44	-0.40	0.51	0.063
st	1.26	0.49	-0.46	0.51	0.039
sr	1.71	0.51	-0.76	0.54	0.005
tr	1.28	0.45	-0.43	0.46	0.035
mst	1.67	0.72	-0.64	0.75	0.047
msr	2.27	0.80	-1.40	0.81	0.002
mtr	1.69	0.67	-1.06	0.68	0.003
str	1.65	0.71	-1.06	0.71	0.004
mstr	2.20	1.02	-1.33	1.05	0.009
dz	1.64	0.43	-0.24	0.48	0.201
dT	2.42	0.76	-1.66	0.81	0.001
df	1.60	0.33	-0.30	0.43	0.085
zT	2.84	0.93	-1.93	0.95	0.001
zf	1.87	0.42	-0.89	0.52	0.002
Tf	2.76	0.79	-1.98	0.85	3E-4
dzT	3.36	1.24	-2.28	1.31	0.002
dzf	2.22	0.63	-1.56	0.80	0.001
dTf	3.27	1.10	-2.31	1.14	0.001
zTf	3.83	1.32	-2.99	1.33	4E-4
dzTf	4.53	1.75	-4.06	1.76	3E-4
msTf	4.76	1.98	-3.68	1.99	0.001

Supplementary Table 6.3 Raw MIC values of Ciprofloxacin (µg/mL) at the end of each day (D)

The values listed here were used to construct graphs for the following strains in Figure 3: *d, z, T, f, msTf,* and *dzTf*.

	d			z			T			f		
Replicate	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
1	0.04	0.04	0.04	0.04	0.08	0.08	0.02	0.04	0.04	0.04	0.04	0.08
2	0.02	0.04	0.08	0.04	0.04	0.04	0.02	0.02	0.04	0.04	0.01	0.04
3	0.04	0.04	0.04	0.04	0.08	0.08	0.01	0.02	0.04	0.04	0.02	0.04
4	0.04	0.04	0.08	0.04	0.04	0.04	0.02	0.02	0.04	0.04	0.08	0.08
5	0.04	0.08	0.08	0.04	0.04	0.08	0.02	0.04	0.08	0.04	0.04	0.04
6	0.04	0.04	0.04	0.04	0.08	0.08	0.02	0.02	0.02	0.04	0.04	0.04
7	0.04	0.04	0.04	0.04	0.04	0.08	0.02	0.04	0.04	0.04	0.04	0.08
8	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.04	0.02	0.04	0.01	0.04
9	0.02	0.04	0.04	0.04	0.04	0.04	0.02	0.04	0.04	0.04	0.04	0.04
10	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.04	0.02	0.04	0.04	0.08
11	0.02	0.02	0.04	0.04	0.04	0.16	0.02	0.02	0.02	0.04	0.04	0.04
12	0.04	0.04	0.04	0.04	0.08	0.08	0.02	0.02	0.02	0.04	0.04	0.04
13	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.04	0.04	0.01	0.01
14	0.04	0.04	0.01	0.04	0.04	0.08	0.02	0.02	0.08	0.04	0.04	0.04
15	0.04	0.08	0.08	0.04	0.04	0.04	0.02	0.04	0.08	0.04	0.04	0.08
16	0.04	0.04	0.08	0.04	0.04	0.04	0.02	0.04	0.08	0.04	0.04	0.04
17	0.02	0.04	0.04	0.04	0.02	0.04	0.02	0.04	0.04	0.04	0.04	0.04
18	0.02	0.04	0.08	0.04	0.04	0.04	0.02	0.04	0.08	0.04	0.04	0.04
19	0.04	0.04	0.04	0.02	0.04	0.08	0.02	0.02	0.02	0.04	0.04	0.04
20	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.04	0.04	0.04	0.16	0.16
21	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.04	0.04	0.04	0.04
22	0.04	0.04	0.01	0.04	0.04	0.08	0.02	0.04	0.04	0.02	0.01	0.04

	C			msTf			dzTf		
Replicate	D1	D2	D3	D1	D2	D3	D1	D2	D3
1	0.04	0.08	0.08	0.04	0.01	0.02	.005	0.02	0.02
2	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.01	.005
3	0.08	0.32	0.32	0.04	0.04	0.04	0.02	0.02	
4	0.04	0.04	0.08	0.02	0.02	0.04			
5	0.04	0.04	0.01	0.08	0.01	0.04	0.01	0.02	
6	0.04	0.04	0.08	0.04	0.01	0.04	0.04	0.02	
7	0.04	0.08	0.08	0.04	0.02	0.02	0.02	0.02	
8	0.04	0.04	0.04	0.02	0.04	0.04	0.04	0.04	0.01
9	0.04	0.02	0.04	0.04	0.01	0.04	0.02		
10	0.04	0.04	0.08	0.02	0.02	0.02	0.01	0.02	
11	0.04	0.02	0.04	0.04	0.01	0.01			
12	0.04	0.04	0.04	0.02	0.01	0.04	0.02	0.04	0.04
13	0.04	0.02	0.04	0.04	0.08	0.01	0.02	0.02	
14	0.04	0.04	0.04	0.04	0.01	0.02	0.04		
15	0.04	0.02	0.08	0.02	0.04	0.04	0.02	0.02	0.02
16	0.04	0.04	0.04	0.04	0.01	0.04	0.02	0.02	
17	0.04	0.04	0.08	0.02	0.01	0.04			
18	0.04	0.04	0.08	0.02	0.02	0.01	0.01	0.01	
19	0.04	0.04	0.04	0.02	0.02	0.01	0.02	0.04	
20	0.04	0.08	0.08	0.02	0.01	0.04	0.04	0.02	0.02
21	0.04	0.04	0.04	0.02	0.02	0.02	0.01		
22	0.04	0.04	0.04	0.02	0.04	0.08	0.04	0.01	0.02

Supplementary Table 6.4 Raw MIC values of Ciprofloxacin (µg/mL) at the end of each day (D)

	m			s			t			r		
Replicate	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
1	0.02	0.02	0.04	0.02	0.02	0.04	0.02	0.04	0.04	0.04	0.08	0.16
2	0.02	0.04	0.08	0.02	0.02	0.04	0.08	0.08	0.08	0.04	0.02	0.08
3	0.02	0.04	0.16	0.02	0.02	0.04	0.02	0.08	0.08	0.04	0.04	0.08
4	0.02	0.04	0.04	0.02	0.04	0.08	0.02	0.04	0.04	0.04	0.08	0.08
5	0.02	0.04	0.04	0.02	0.08	0.08	0.04	0.08	0.16	0.04	0.04	0.08
6	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.04	0.04	0.08	0.08	0.08
7	0.02	0.04	0.04	0.02	0.04	0.04	0.08	0.08	0.08	0.04	0.08	0.08
8	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.04	0.08	0.02	0.04	0.08
9	0.02	0.16	0.32	0.02	0.04	0.16	0.02	0.02	0.04	0.02	0.04	0.04
10	0.04	0.08	0.08	0.04	0.04	0.04	0.02	0.08	0.08	0.04	0.02	0.04
11	0.02	0.04	0.08	0.02	0.16	0.16	0.02	0.04	0.04	0.04	0.08	0.16
12	0.02	0.04	0.04	0.02	0.08	0.08	0.02	0.04	0.02	0.04	0.08	0.08
13	0.02	0.04	0.08	0.04	0.04	0.08	0.02	0.04	0.08	0.04	0.04	0.08
14	0.04	0.08	0.08	0.04	0.04	0.08	0.02	0.08	0.08	0.04	0.04	0.08
15	0.02	0.04	0.04	0.02	0.08	0.08	0.02	0.04	0.04	0.04	0.08	0.08
16	0.02	0.04	0.08	0.02	0.04	0.04	0.02	0.04	0.04	0.04	0.04	0.08
17	0.02	0.04	0.04	0.02	0.04	0.08	0.02	0.02	0.04	0.04	0.02	0.04
18	0.08	0.08	0.08	0.02	0.02	0.04	0.08	0.08	0.16	0.02	0.04	0.08
19	0.02	0.04	0.04	0.02	0.04	0.08	0.02	0.08	0.16	0.04	0.04	0.08
20	0.02	0.04	0.08	0.02	0.04	0.04	0.02	0.04	0.08	0.04	0.04	0.08
21	0.02	0.08	0.08	0.02	0.02	0.04	0.02	0.08	0.08	0.04	0.04	0.08
22	0.04	0.08	0.08	0.02	0.04	0.02	0.02	0.04	0.02	0.04	0.04	0.08

	C			CCCC			mstr		
Replicate	D1	D2	D3	D1	D2	D3	D1	D2	D3
1	0.02	0.04	0.08	0.08	0.08	0.08	0.02	0.02	0.02
2	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.08	0.08
3	0.02	0.08	0.08	0.02	0.04	0.02	0.04		
4	0.02	0.04	0.02	0.04	0.08	0.08	0.02	0.04	0.01
5	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.04	0.02
6	0.02	0.04	0.04	0.04	0.08	0.08	0.08	0.04	
7	0.04	0.08	0.08	0.04	0.04	0.04	0.02	0.04	0.04
8	0.02	0.04	0.04	0.02	0.04	0.04	0.08	0.08	0.08
9	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.02	0.02
10	0.02	0.04	0.02	0.02	0.04	0.04	0.02	0.04	0.04
11	0.02	0.04	0.04	0.02	0.04	0.04	0.04	0.08	0.08
12	0.02	0.04	0.04	0.04	0.04	0.04	0.02	0.04	0.08
13	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.04	0.02
14	0.04	0.08	0.16	0.02	0.04	0.04	0.02	0.04	
15	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.04	0.04
16	0.02	0.04	0.04	0.02	0.04	0.02	0.02	0.04	0.04
17	0.04	0.04	0.08	0.04	0.08	0.08	0.02	0.04	0.08
18	0.02	0.04	0.04	0.02	0.04	0.04	0.02	0.08	0.02
19	0.04	0.08	0.04	0.02	0.04	0.04	0.02	0.04	0.16
20	0.02	0.02	0.04	0.04	0.04	0.04	0.08	0.08	0.02
21	0.04	0.04	0.04	0.02	0.04	0.08	0.08	0.08	0.16
22	0.02	0.04	0.08	0.04	0.04	0.08	0.02	0.04	0.08

Supplementary Table 6.5 Raw MIC values of Ciprofloxacin (µg/mL) from a separate repeat experiment testing dzTf

	C			dzTf		
Replicate	D1	D2	D3	D1	D2	D3
1	0.04	0.04	0.04	0.02	0.04	.005
2	0.04	0.04	0.04	0.04	0.01	.005
3	0.04	0.04	0.04	0.04	0.02	
4	0.04	0.04	0.16	0.04	0.04	0.04
5	0.04	0.08	0.08			
6	0.04	0.04	0.04	0.04	0.04	0.04
7	0.04	0.04	0.04	0.04	0.04	0.04
8	0.04	0.08	0.08	0.04	0.04	0.04
9	0.04	0.08	0.16			
10	0.04	0.04	0.04	.005	0.04	
11	0.04	0.04	0.04			
12	0.04	0.08	0.16	0.04	0.04	0.01
13	0.04	0.04	0.04	0.04	.005	0.01
14	0.04	0.08	0.16	0.04	0.02	
15	0.04	0.04	0.04			
16	0.04	0.08	0.04	0.04	0.04	
17	0.04	0.04	0.08	0.04	0.04	0.01
18	0.04	0.04	0.04	0.04	0.02	
19	0.04	0.04	0.08	0.04	0.02	
20	0.04	0.04	0.04	0.02	0.04	
21	0.04	0.04	0.04			
22	0.04	0.04	0.04	0.04	0.02	

Supplementary Table 6.6 Sequencing of gyrA after 3 days of cipro. adaptation

Strain	Replicate	Day 3 MIC (µg/mL)	Mutation
C	1	0.08	N/A
			N/A
	3	0.32	Point mutation – D87Y
			Point mutation – D87Y
	4	0.08	N/A
			N/A
d	2	0.08	N/A
			N/A
	4	0.08	N/A
			N/A
	5	0.08	N/A
			N/A
z	1	0.08	N/A
			N/A
	3	0.08	N/A
			N/A
	11	0.16	N/A
			N/A
T	5	0.08	N/A
			N/A
	14	0.08	Codon deletion – S83
			Codon deletion – S83
	15	0.08	N/A
			N/A
f	1	0.08	N/A
			N/A
	4	0.08	N/A
			N/A
	20	0.16	N/A
			N/A
msTf	3	0.04	N/A
			N/A
	4	0.04	N/A
			N/A
	22	0.08	N/A
			N/A
dzTf	1	0.02	N/A
			N/A
	12	0.04	N/A
			N/A
	20	0.02	N/A
			N/A

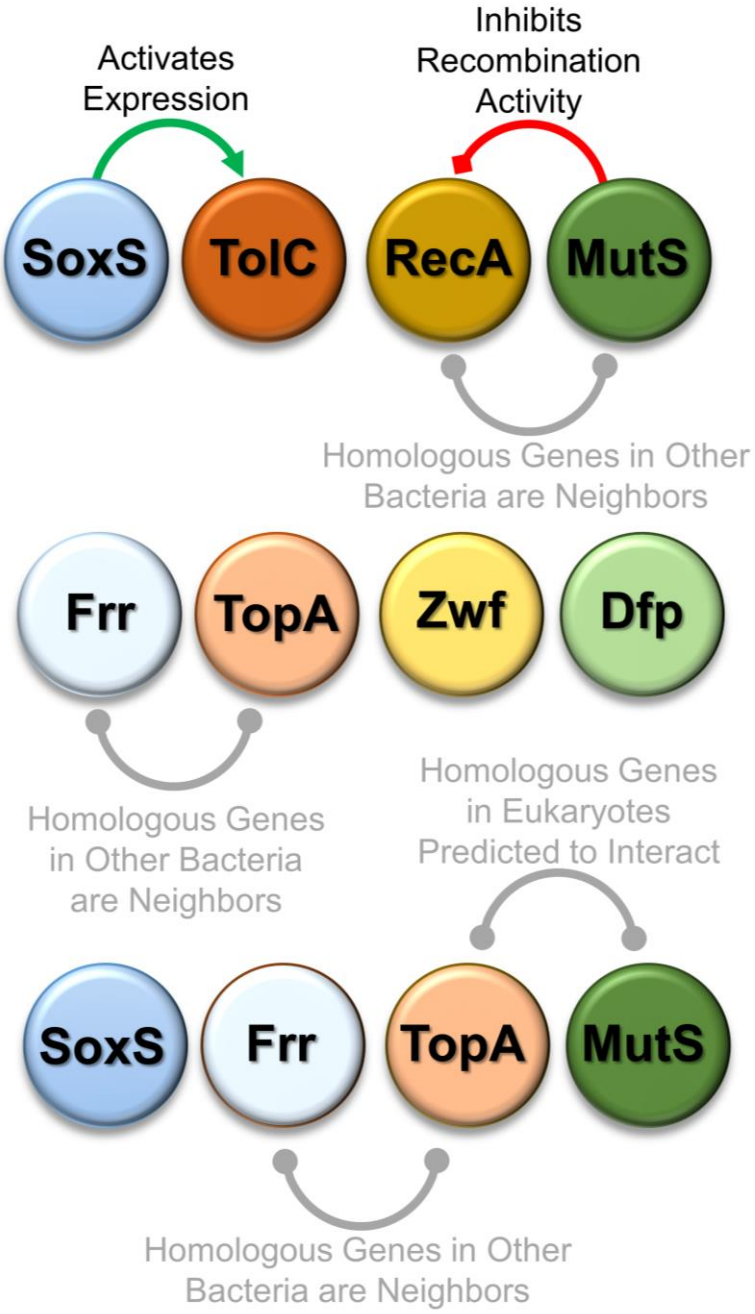
Supplementary Table 6. Continued

Strain	Replicate	Day 3 MIC (µg/mL)	Mutation
C	2	0.08	N/A
			N/A
	4	0.08	N/A
			N/A
	15	0.16	N/A
			N/A
CCCC	2	0.08	N/A
			N/A
	5	0.08	N/A
			N/A
	7	0.08	N/A
			N/A
m	3	0.08	N/A
			N/A
	4	0.16	N/A
			N/A
	10	0.32	N/A
			N/A
s	5	0.08	N/A
			N/A
	10	0.16	N/A
			N/A
	12	0.16	Point mutation – D82G
			Point mutation – D82G
t	6	0.16	N/A
			N/A
	19	0.16	Codon deletion – S83
			N/A
	20	0.16	N/A
			N/A
r	2	0.16	N/A
			N/A
	3	0.08	N/A
			N/A
	12	0.16	N/A
			N/A
mstr	3	0.08	N/A
			N/A
	20	0.16	N/A
			N/A
	22	0.16	N/A
			N/A

Supplementary Table 6.7 Cloning, sequencing, and RT-qPCR primers used in this study. Target locations of sgRNAs are highlighted in red

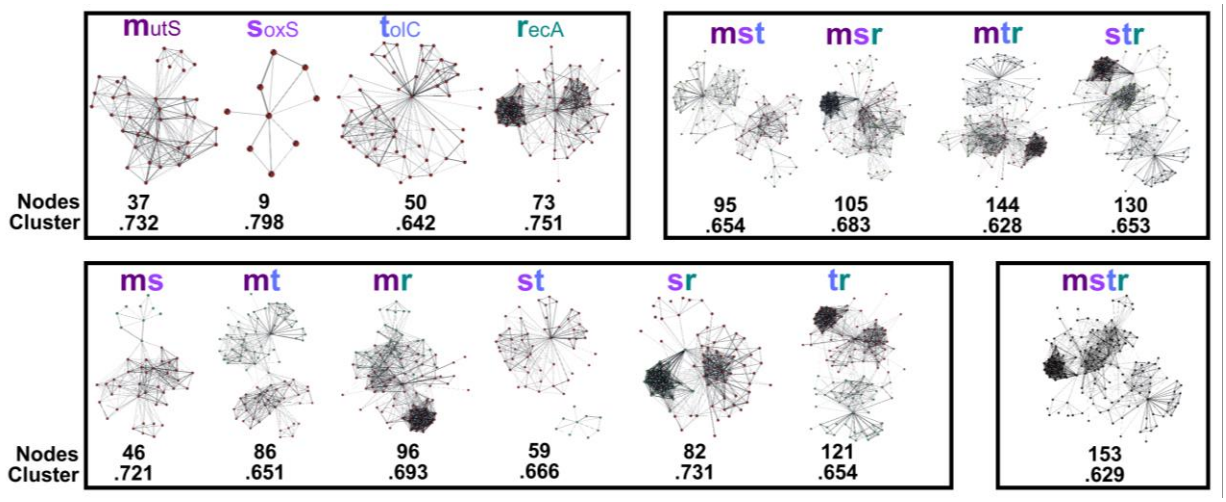
Primer Name	Sequence (5'-3')
sgRNA cloning forward – *mutS* activation	ACTAGTACTAGT**GCAAGTACGCAAAATTGTAT**GTTTTAGAGCTAGAAATAGC
sgRNA cloning forward – *soxS* activation	ACTAGTACTAGT**GCGTTTCGCCACTTCGCCGG**GTTTTAGAGCTAGAAATAGC
sgRNA cloning forward – *tolC* activation	ACTAGTACTAGT**AGCAGTCATGTGTTAAATTG**GTTTTAGAGCTAGAAATAGC
sgRNA cloning forward – *recA* activation	ACTAGTACTAGT**CCGTGATGCGGTGCGTCGTC**GTTTTAGAGCTAGAAATAGC
sgRNA cloning forward – *dfp* inhibition	ACTAGTACTAGT**GTGATAAAATCGCCAACTTC**GTTTTAGAGCTAGAAATAGC
sgRNA cloning forward – *zwf* inhibition	ACTAGTACTAGT**GTATACTTGTAATTTTCTTA**GTTTTAGAGCTAGAAATAGC
sgRNA cloning forward – *topA* inhibition	ACTAGTACTAGT**CTGGCAACGAGTTACCGATA**GTTTTAGAGCTAGAAATAGC
sgRNA cloning forward – *frr* inhibition	ACTAGTACTAGT**AGCCCTGATTAAACATATTA**GTTTTAGAGCTAGAAATAGC
sgRNA cloning single target reverse	GGGCCCGGGCCCAAGCTTCAAAAAAAGCACCG
sgRNA plasmid sequencing	GGGGGGGACGTCTAAGAAACCATTATTATCATG
Gibson primer – backbone Forward	CTCAGGTTCATCGTTTGCATGGGATCACAGGTCTAAGAAACCATTATTATCATGACATTA
Gibson primer – backbone Reverse	ATTGACAGCTGAGAGCGCCTTCACAACGTGGTCAGGTGGCACTTTTCGGGGAAATGTGCG
Gibson primer – insert 1 Forward	CACGTTGTGAAGGCGCTCTCAGCTGTCAATCTCGAGTAAGGATCCAGTTCACCGACAAAC
Gibson primer – insert 1 Reverse	CTGTGATCCCATGCAAACGATGAACCTGAGAATTCTAAAGATCTTTGACAGCTAGCTCAG
Gibson primer – insert 2 Forward	TCACCAAATTATAGCCATACAGACCCAAATCTCGAGTAAGGATCCAGTTCACCGACAAAC
Gibson primer – insert 2 Reverse	ATTTGGGTCTGTATGGCTATAATTTGGTGAAATTCTAAAGATCTTTGACAGCTAGCTCAG
Gibson primer – insert 3 Forward	TTTAAATCATATCACACAATTAGCCTCTCGCTCGAGTAAGGATCCAGTTCACCGACAAAC
Gibson primer – insert 3 Reverse	CGAGAGGCTAATTGTGTGATATGATTTAAAAATTCTAAAGATCTTTGACAGCTAGCTCAG
Cloning of mCherry from plasmid PHL662 to sgRNA plasmid, Forward	GACGTCGACGTCTAAGAAACCATTATTATCAT
Cloning of mCherry from plasmid PHL662 to sgRNA plasmid, Reverse	GACGTCGACGTCTTACTTGTACAGCTCGTCCA
gyrA sequencing amplification Forward	GCCTTCCACGCGTTTTTCTT
gyrA sequencing amplification Reverse	AACCGACATCGAGCACCTTT
qPCR *mutS* Forward	ATGGAACGTGAGCAGGACAG
qPCR *mutS* Reverse	CAGCCAGCGTTTCAGCATAC
qPCR *soxS* Forward	TCTGCTGCGAGACATAACCC
qPCR *soxS* Reverse	ACTTGCAACGAATGTTCCGC
qPCR *tolC* Forward	ACGCACTACCACCAGTAACG
qPCR *tolC* Reverse	TTTGTCTTCCGGGACCAGTG
qPCR *recA* Forward	ATCGCCTGGCTCATCATACG
qPCR *recA* Reverse	GCACTGGAAATCTGTGACGC
qPCR *dfp* Forward	TTTGATTGCCCGTGTTGCTG
qPCR *dfp* Reverse	CGGGAAGCAAGCACCTCTAA
qPCR *zwf* Forward	ACGAAGTGGAAGAAGCCTGG
qPCR *zwf* Reverse	TCACGGGTAATCATCGCCAC
qPCR *topA* Forward	CTTGTGGTCGCAAAATGGGG
qPCR *topA* Reverse	GCACGTTCAGCACTTCGTTT
qPCR *frr* Forward	AAGCGATTATGGCGTCCGAT
qPCR *frr* Reverse	GCTTGTTCTGCTTCACCACG
qPCR *gyrA* Forward	GTCATAGACCGCCGAGTCAC
qPCR *gyrA* Reverse & *gyrA* sequencing	GCGATGTCGGTCATTGTTGG
qPCR *cysG* Forward	ATTCCGTTCTCGGTGGTTCC
qPCR *cysG* Reverse	CCAGCGTCTGTTTTTCTGCC
qPCR *dCas9* Forward	AGATTCGCAAACGCCCTCTA
qPCR *dCas9* Reverse	CTCCTTGGAGAATCCGCCTG

6.8.2 Supplementary Figures



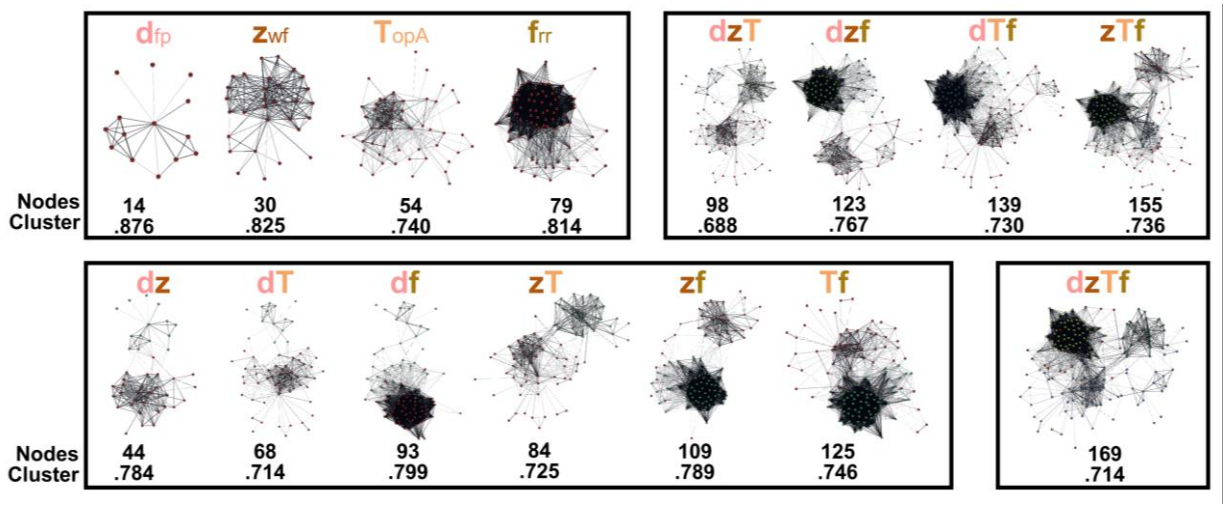
Supplementary Figure 6.1 All known genetic interactions of gene targets investigated in this study

Possible uncharacterized interactions are also listed in gray, based on homology to other organisms. Interactions were determined from the STRING protein database.



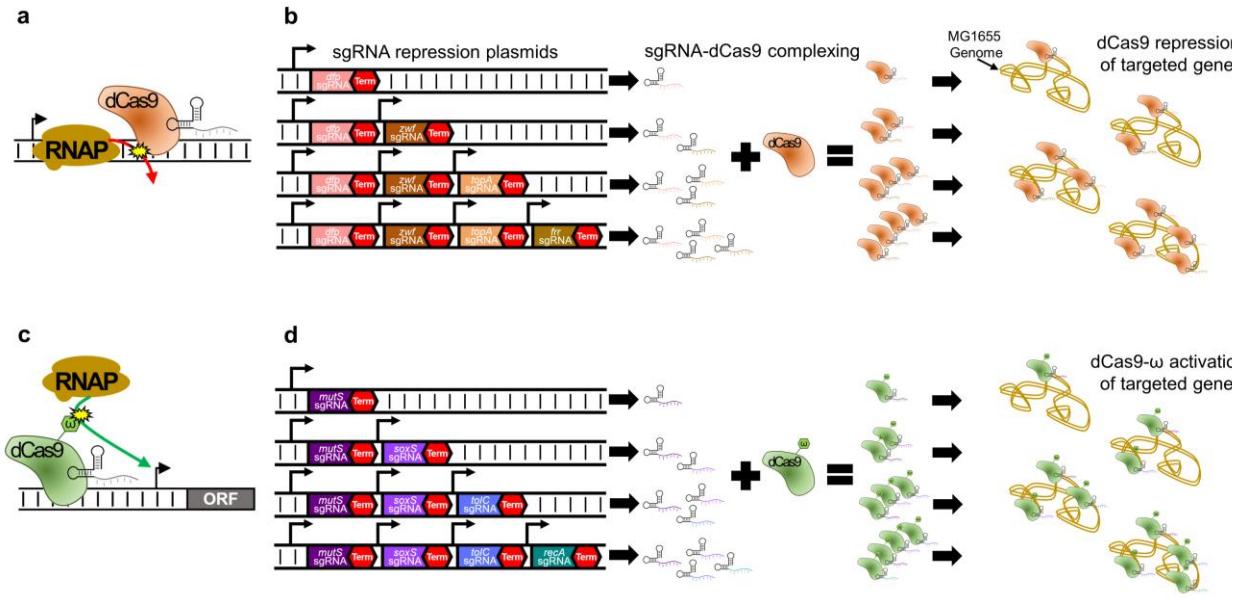
Supplementary Figure 6.2 Predicted protein networks

Networks show proteins impacted by each gene perturbation within the stress response set, including all combinations of perturbations affecting *mutS*, *soxS*, *tolC,* and *recA* abbreviated as in Figure 2. All known potential primary protein partners were determined from the STRING protein database using information from databases, co-expression, co-occurrence, gene fusion, and experimental validation. A minimum interaction score confidence level of 0.4 was used. Each node represents a unique protein, with lines between each node representing a predicted interaction. Line thickness represents the predicted confidence of protein-protein interaction. Below each network is listed the number of nodes contained within, as well as the cluster coefficient. Cluster coefficients predict the overall interconnectedness of the network, with higher values indicating a “tighter” network of more closely associated interactions.



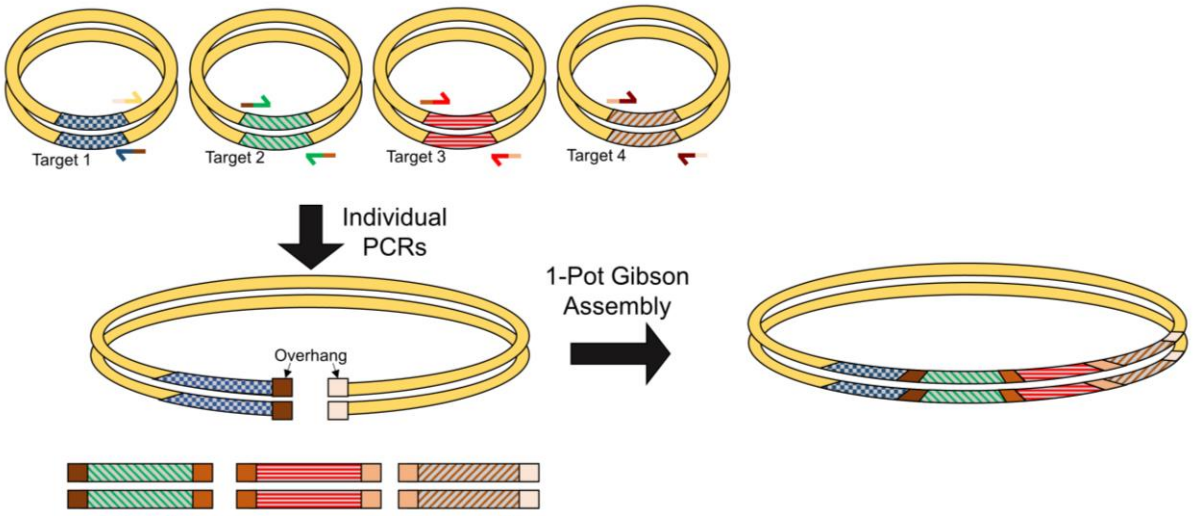
Supplementary Figure 6.3 Predicted protein networks

Networks show proteins impacted by each gene perturbation within the conserved set, including all combinations of perturbations affecting *dfp*, *zwf*, *topA* and *frr* abbreviated as in Figure 2. All known potential primary protein partners were determined from the STRING protein database[280] using information from databases, co-expression, co-occurrence, gene fusion, and experimental validation. A minimum interaction score confidence level of 0.4 was used. Each node represents a unique protein, with lines between each node representing a predicted interaction. Line thickness represents the predicted confidence of protein-protein interaction. Below each network is listed the number of nodes contained within, as well as the cluster coefficient. Cluster coefficients predict the overall interconnectedness of the network, with higher values indicating a "tighter" network of more closely associated interactions.



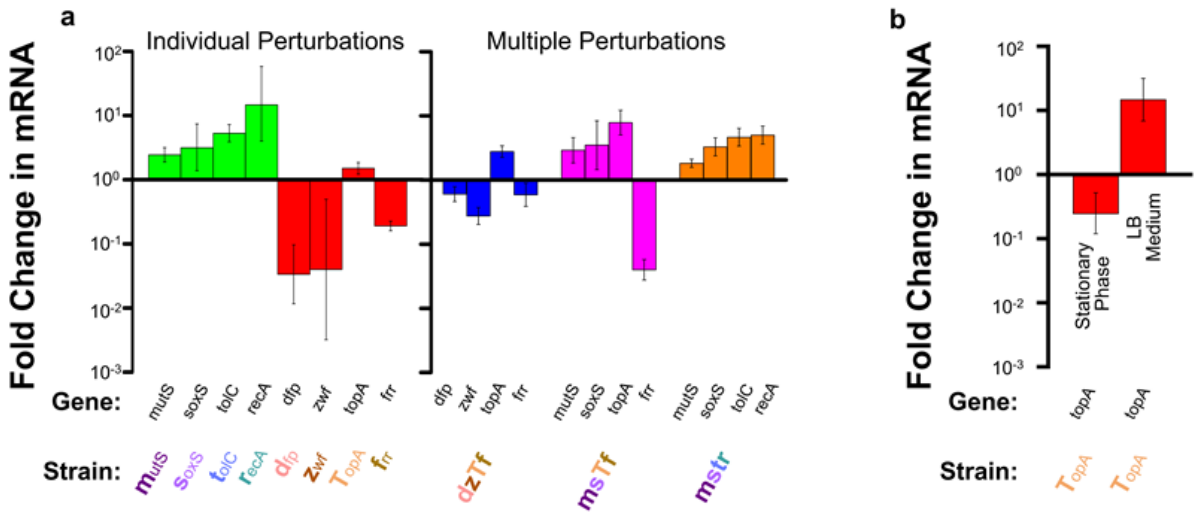
Supplementary Figure 6.4 Schematic depicting how dCas9 and dCas9-ω were used to perturb genes for CHAOS therapy

(**a**) dCas9 is designed to bind the +1 transcription start site of the gene targeted for expression inhibition. Upon dCas9 binding to DNA, a roadblock to transcription is created which prevents read-through of RNA polymerase (RNAP), terminating transcription and reducing mRNA production (and therefore overall gene expression). (**b**) sgRNA plasmids were created in tandem using the approach presented in Figure S5 on a separate plasmid from dCas9. These sgRNAs are independently expressed to create unique guides for particular locations of DNA. Upon complexing with dCas9, these sgRNAs guide the protein-RNA complex to unique locations on the MG1655 chromosome to inhibit expression. (**c**) dCas9-ω is designed to bind 80-100 nt upstream of the +1 promoter site for the targeted gene to activate expression. The ω-subunit of RNAP that is linked to dCas9 serves to recruit RNAP complexes to the +1 promoter site. This increases the copies of RNAP producing mRNA and thereby increases total gene expression. (**d**) In a very similar fashion as sgRNAs designed for complexing with dCas9, sgRNAs designed for complexing with dCas9-ω recruit the protein-RNA complex to unique locations for gene expression activation. As has been demonstrated in our previous work, dCas9-ω can serve in place of dCas9 to inhibit expression, and was used for inhibiting *topA*/*frr* while activating *mutS*/*soxS* in strain “msTf”.



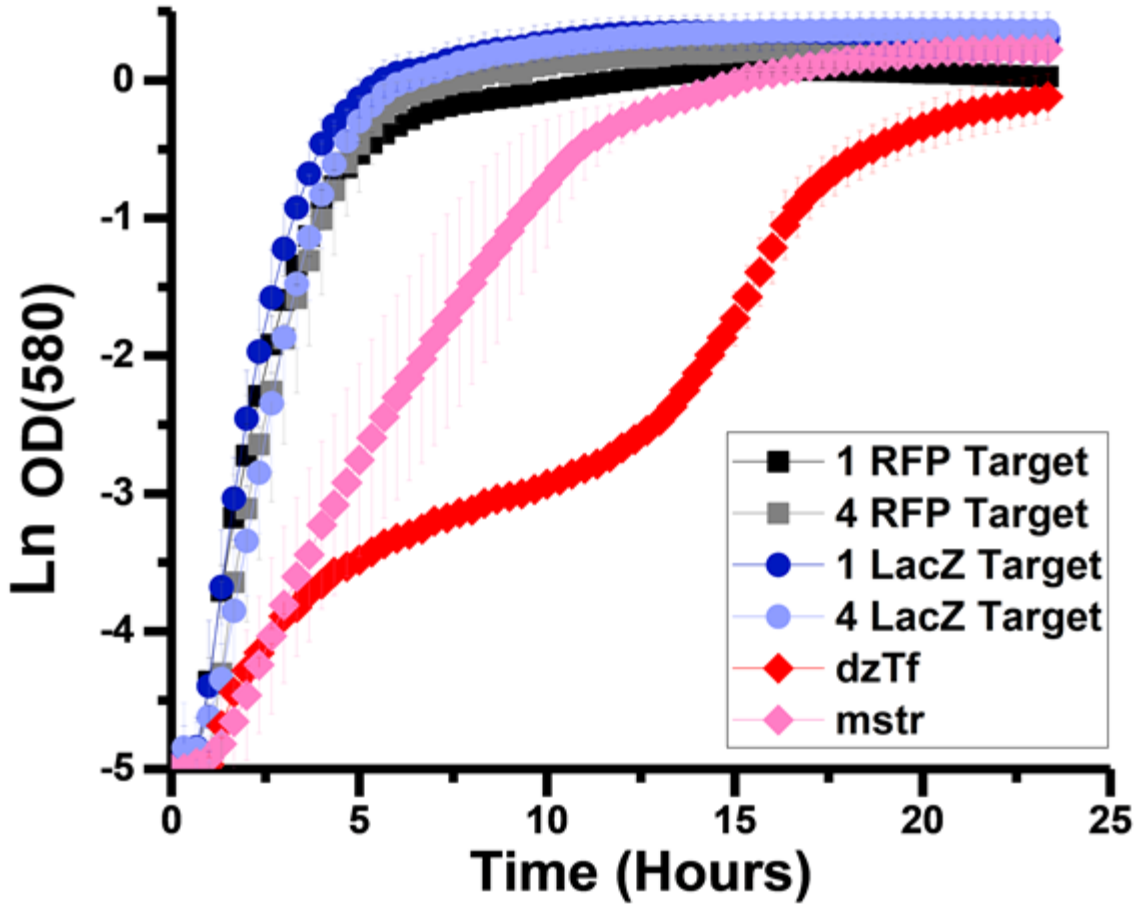
Supplementary Figure 6.5 Schematic depicting strategy for assembling multiple-targeting sgRNA plasmids

Individual targeting sgRNAs were first assembled, from which Gibson Assembly primers were used to amplify targets for Gibson Assembly as depicted. Sizes of final plasmids were confirmed using gel electrophoresis before transferring into the final experimental strain.



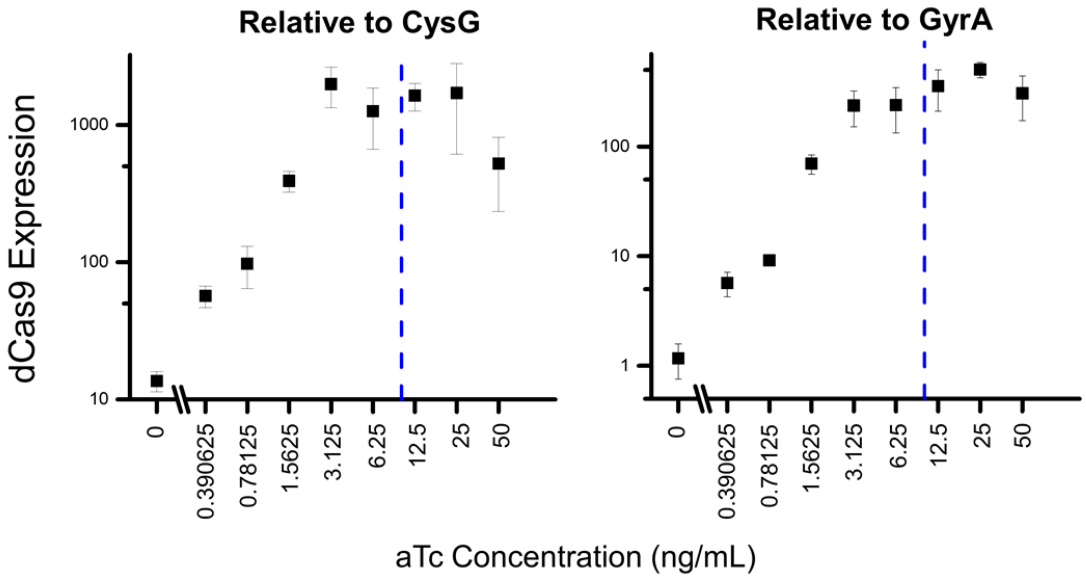
Supplementary Figure 6.6 RT-qPCR results of CRISPR perturbation on gene expression, as quantified by changes in mRNA concentration in relation to the single *rfp* control strain

(**a**) All qPCR was performed under 10 ng/mL aTc induction in M9 minimal media. All samples were collected after 8 hours of growth after induction during exponential phase. All error bars represent standard deviations of biological triplicates, with the average Ct of technical duplicates used for each replicate. The authors note that *recX* expression was likely increased in the strain activating *recA*, although such an increase would be substantially less than that experienced by *recA* due to a strong terminator sequence located in-between *recA* and *recX*. (**b**) We further explored the impact of CRISPR inhibition of *topA*, as this actually resulted in gene activation. We attribute this result to the inherent complex regulation of *topA*, which can be expressed from five unique promoters and is regulated by prolific binding of the phase-dependent and media-dependent transcriptional regulator, Fis. Binding of dCas9 to this promoter region, which was intended to exclude binding of RNA polymerase to the promoter closest to *topA*'s ORF, possibly instead precluded Fis binding in this region. We therefore tested *topA* expression relative to the Control strain in strains grown in both stationary phase in M9 minimal media, and in exponential phase in LB medium. We found that in stationary phase, where Fis expression is significantly lower, CRISPR inhibition worked as intended. Conversely, in LB medium, where Fis expression is significantly higher than in minimal medium, the unintended activation of *topA* by the CRISPR inhibition construct was exacerbated. These results strongly imply that the degree of *topA* perturbation by the CRISPR construct was heavily dependent on Fis concentration in the cell.



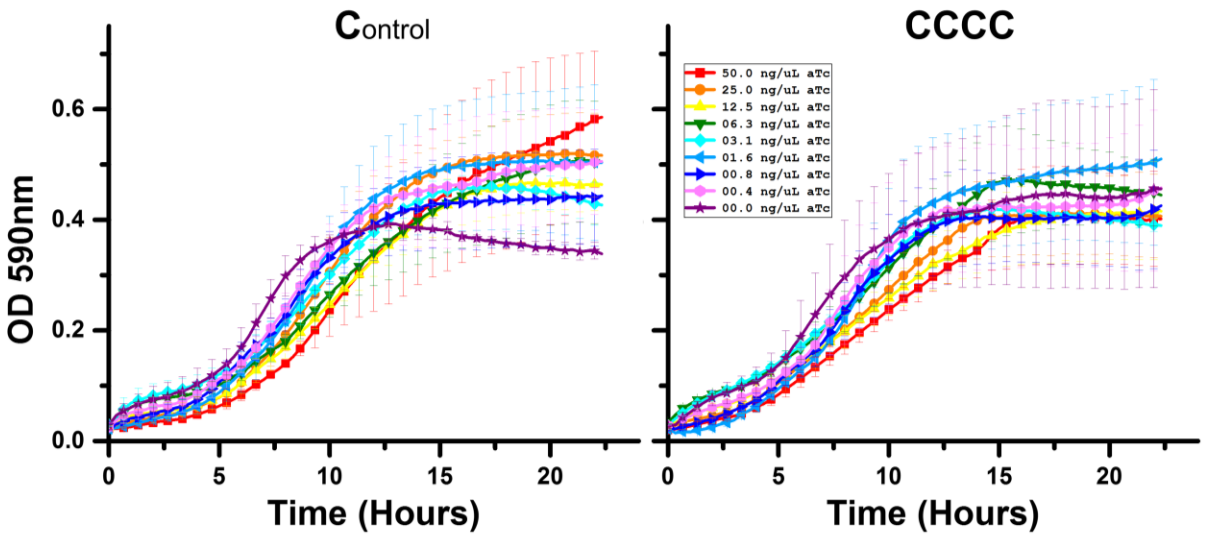
Supplementary Figure 6.7 LacZ Growth Curves

Growth curves of CRISPR perturbation strains harboring one or four copies of control perturbations inhibiting expression of RFP (whose target sequence is absent from said strain), or inhibiting expression of lacZ (whose expression is not expected to affect bacterial fitness in LB medium). Also included are growth curves of the four perturbation strains targeting all four conserved genes (dzTf) or all four stress response genes (mstr) to demonstrate that selective pressure was indeed present. All growth was performed in LB medium supplemented with 10 ng/mL aTc and 0.005 µg/mL ciprofloxacin. Error bars represent standard deviation of five biological replicates.



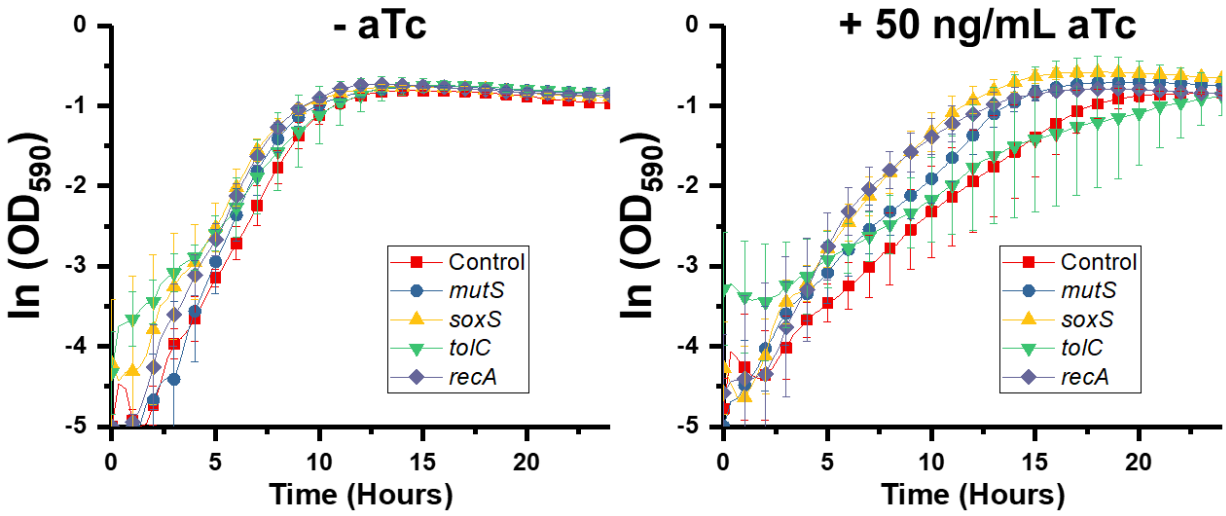
Supplementary Figure 6.8 Quantifying dCas9 leaky expression

Relative expression of dCas9 at various levels of aTc induction, as determined by RT-qPCR. Notably, significant leaky expression of dCas9 was observed even under no induction. Saturation of aTc induction appeared at ~3 ng/mL. The concentration of aTc used in experiments in Figs. 2-4 was 10 ng/mL and is indicated by blue dashed lines. This concentration induced ~100 fold more dCas9 expression than the basal leaky expression during inclusion of no aTc.



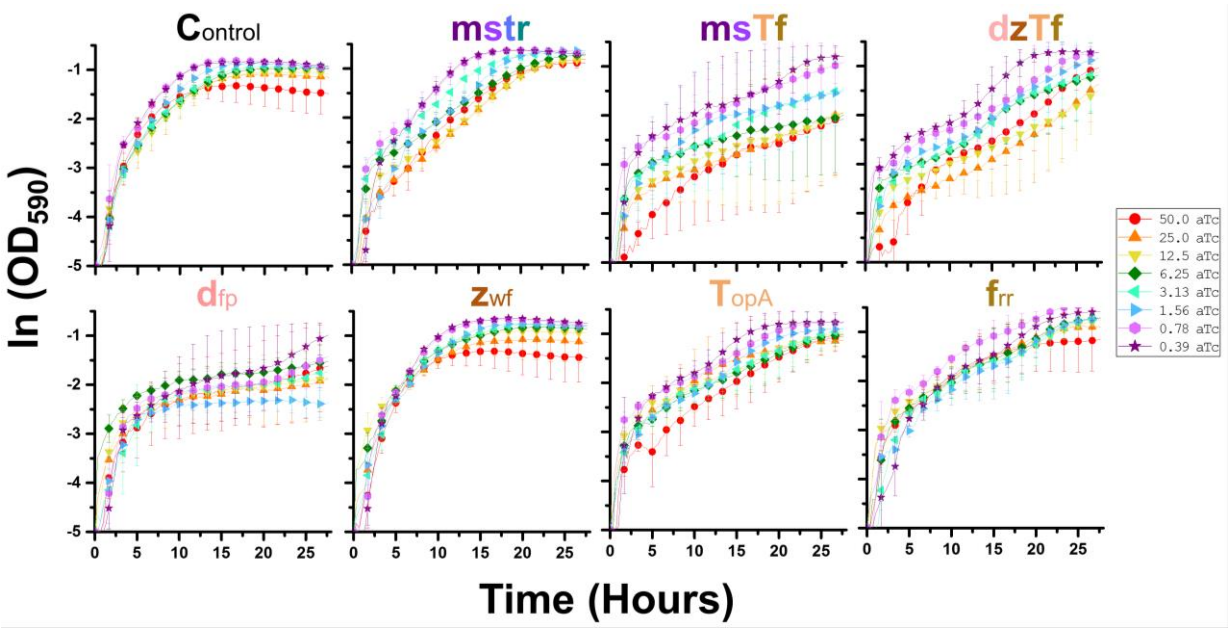
Supplementary Figure 6.9 Control growth curves

Growth curves of control strains under various concentrations of aTc, representing various levels of induction of dCas9 and the CRISPR perturbation system at large. A slight growth deficit was observed from aTc toxicity correlating with higher concentrations, as would be expected for a strain of *E. coli* harboring two plasmids grown in minimal media maintaining two-antibiotic selection pressure. Error bars represent standard deviation of biological triplicates.



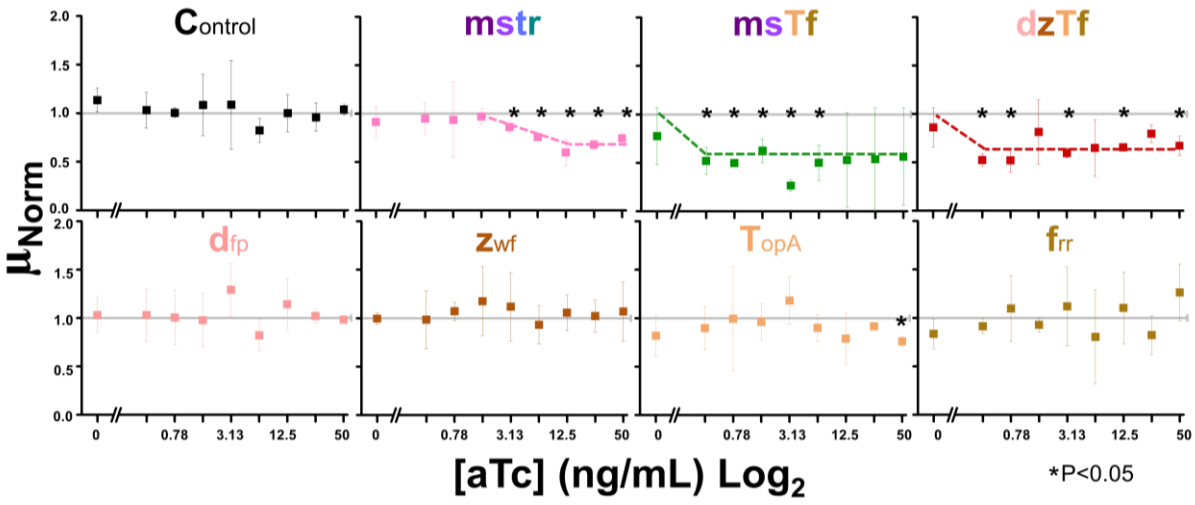
Supplementary Figure 6.10 Stress perturbation growth curves

Growth curves of individual perturbations of the four stress response genes with and without aTc. Optical densities are converted to logarithmic form and normalized to the starting value to highlight the exponential phase of growth.



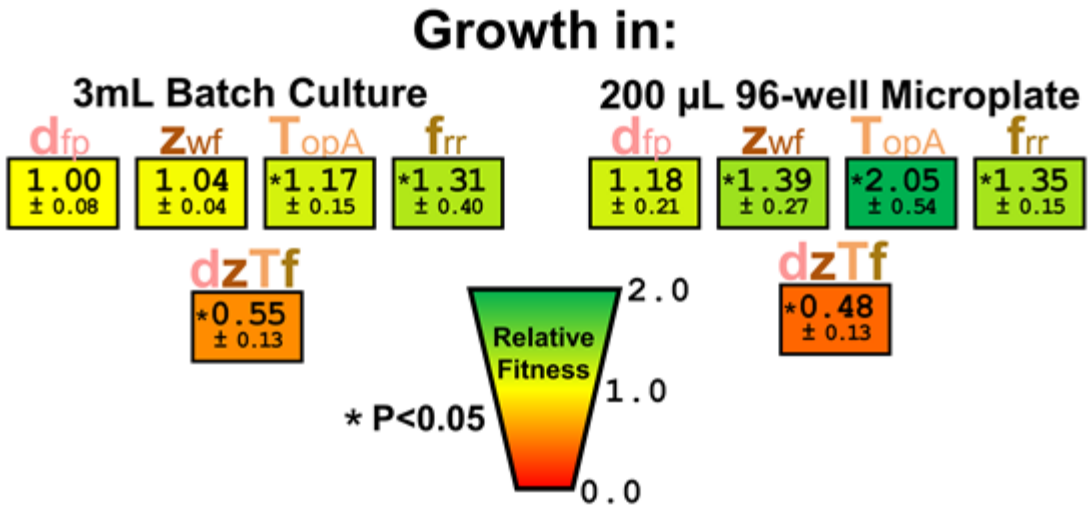
Supplementary Figure 6.11 Growth curves of conserved and multiplexed perturbations

Growth curves of four perturbation strains under various levels of induction of the CHAOS CRISPR perturbation system. Optical densities are converted to logarithmic form and normalized to the starting value to highlight the exponential phase of growth. The concentration of aTc used to induce expression (ng/mL) is indicated in the legend.



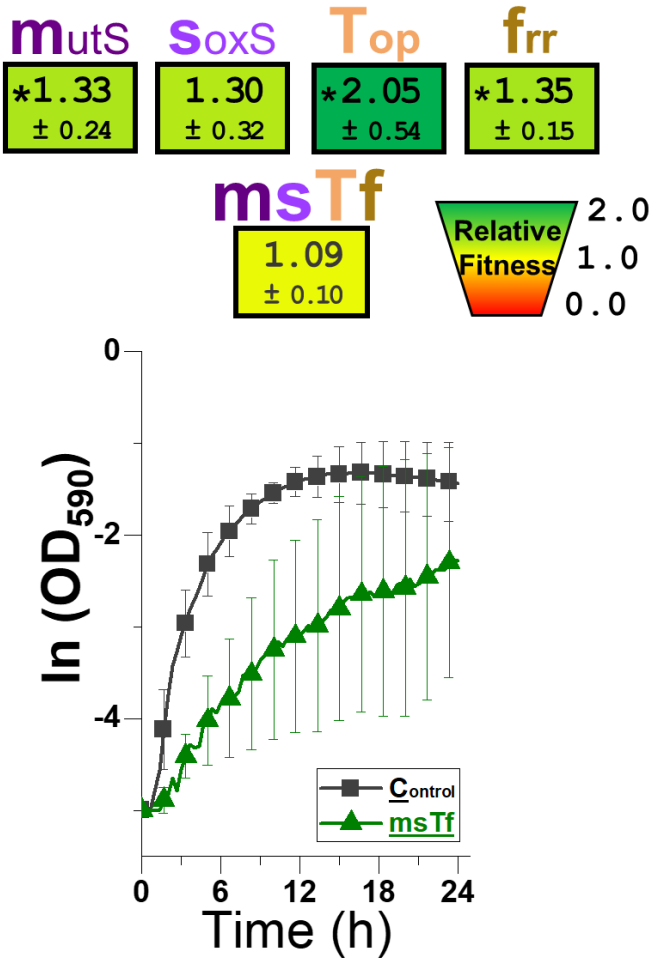
Supplementary Figure 6.12 Growth rates at various induction levels

Normalized growth rates of select strains under different levels of induction of the CHAOS CRISPR perturbation system. Growth rates are normalized to the single *rfp* target control strain grown under the same concentration of aTc. Dashed lines are included for the visual aid of trends. Asterisks indicate statistically significant differences in normalized growth rates from the control (P-value <0.05, two-tailed type II t-test).



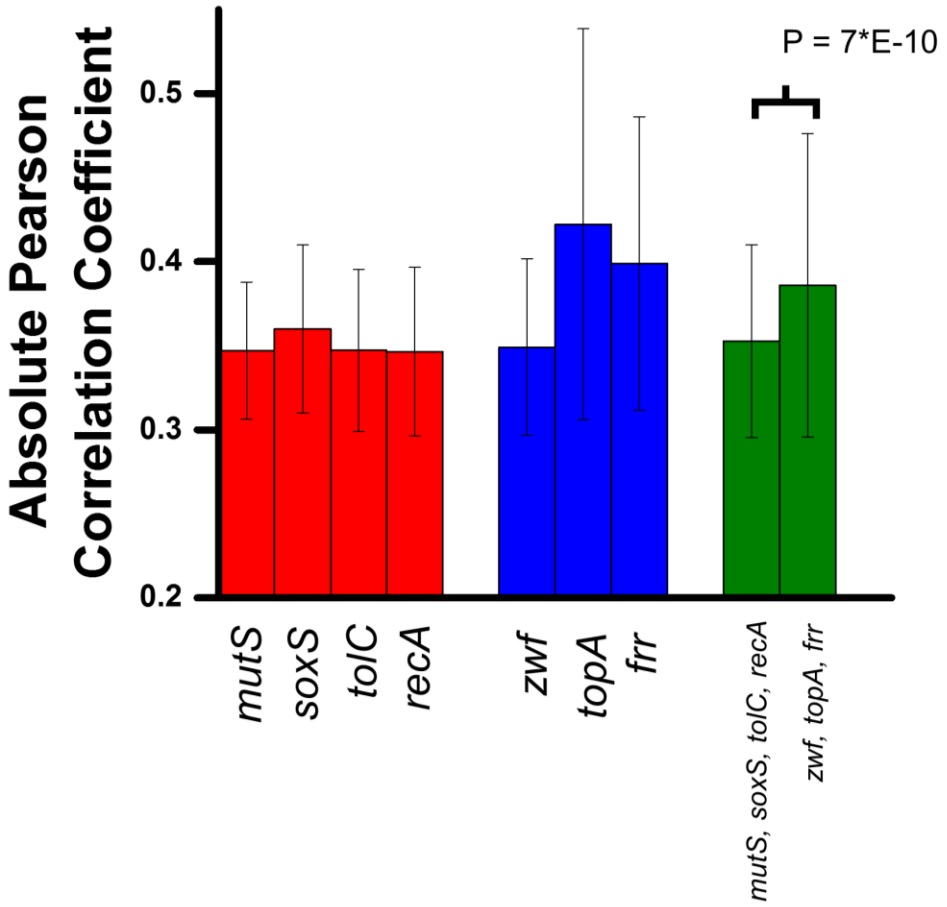
Supplementary Figure 6.13 Exploration of the impact of growth in microplates on experimental outcomes

Exploration of the impact of growth in microplates on experimental outcomes. The four individual perturbations of conserved genes, as well as the combined perturbations of all four simultaneously, were competed again against $C_{mCherry}$ during growth in LB medium supplemented with 10 ng/mL aTc and 0.005 µg/mL ciprofloxacin. However, growth was performed instead in 3mL LB in 5mL glass culture tubes with continuous shaking to improve culture aeration. Competitive fitness under these conditions is presented on the left, with original microplate-based fitness presented on the right. Overall, a slight reduction in the fitness benefits or detriments of perturbations appeared to emerge during growth in batch culture, likely due to an increase in cell's health in these more favorable growth conditions. However, the significant trend of lower fitness as perturbations were combined remained, suggesting that culture growth in microplates does not impact the overall finding of emergent epistasis at the gene expression level. Standard deviations were calculated from three biological triplicates for growth in 3mL batch cultures. Asterisks indicate significant differences from the eight biological replicates of strain C_{ontrol} grown in the same conditions as the right.



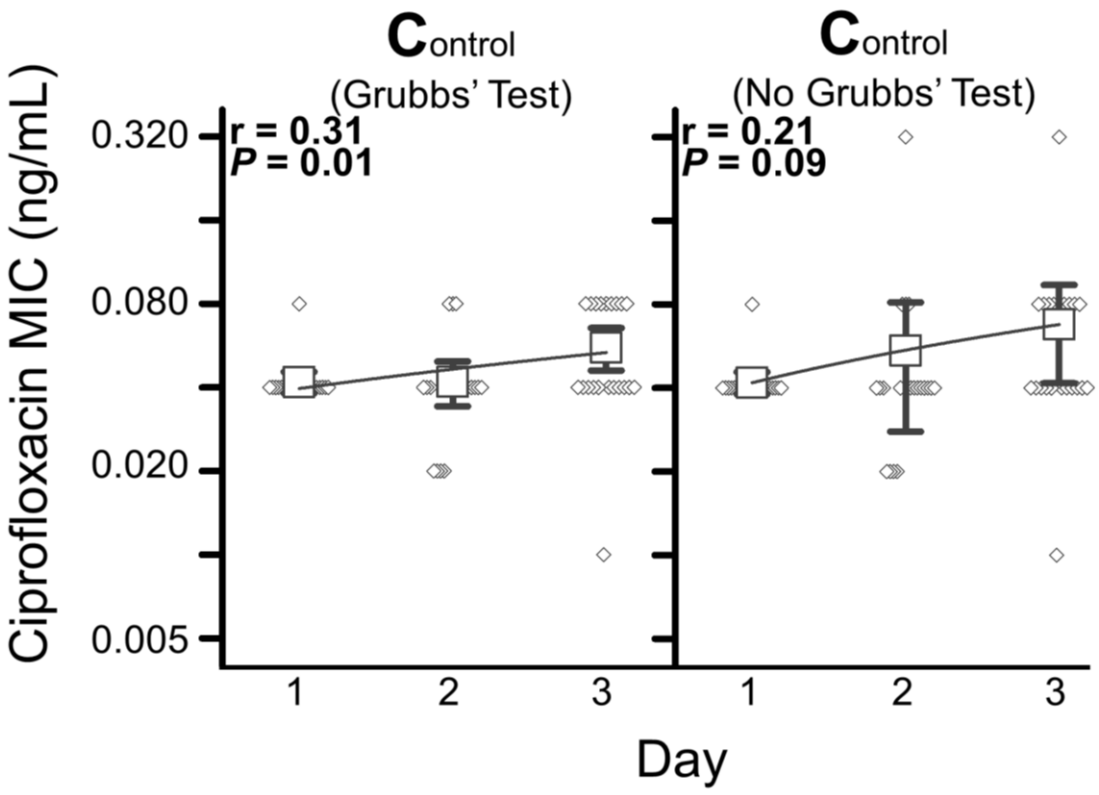
Supplementary Figure 6.14 msTf fitness

Fitness of strains harboring either CRISPR activations of stress response genes (*mutS* and *soxS*) or CRISPR inhibitions of conserved genes (*topA* and *frr*) during exposure to 0.005 µg/mL ciprofloxacin and 10 ng/mL aTc in LB medium. A strain activating expression of the first two while inhibiting expression of the latter two was also constructed and tested in the same fashion. Relative fitness is listed below each strain name, followed by the standard deviation (n = 8). Asterisks indicate significant fitness differences in relation to strain "C_{ontrol}" from Figure 2 ($P < 0.01$, two-tailed type II *t*-test). Additionally, growth of strain msTf relative to the control strain was performed during exposure to 50 ng/mL aTc in M9 minimal media.



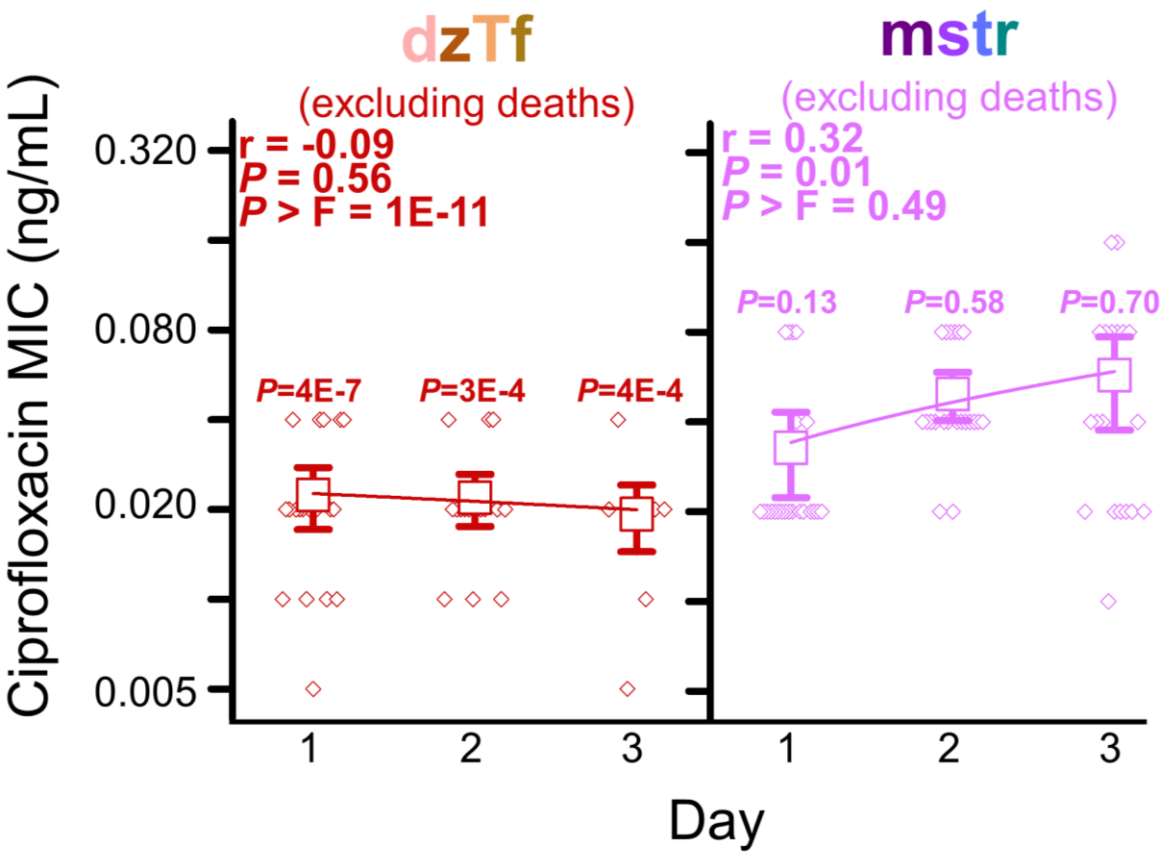
Supplementary Figure 6.15 Genetic interactions of mstr and zTf

Average absolute Pearson Correlation Coefficient (PCC) of statistically significant genetic interactions determined by Babu et al[286]. In this study, approximately 600,000 double-mutant strains were created from 163 gene knockouts crossed with 3,968 non-essential single gene deletions and 149 hypomorphic mutations. PCCs of phenotypes were calculated for every gene against all other genes, depicting the relative strength of that gene's genetic interaction with other genes. We took the average of the absolute value of each reported PCC of both individual genes, as well as the average within each set of genes investigated. This study did not investigate *dfp*, hence its exclusion here. We note that the average value of PCC of *mutS*, *soxS*, *tolC*, and *recA* is less than the average PCC of *zwf*, *topA*, and *frr* suggesting that the genetic interactions of the former with other genes throughout the genome are typically weaker.



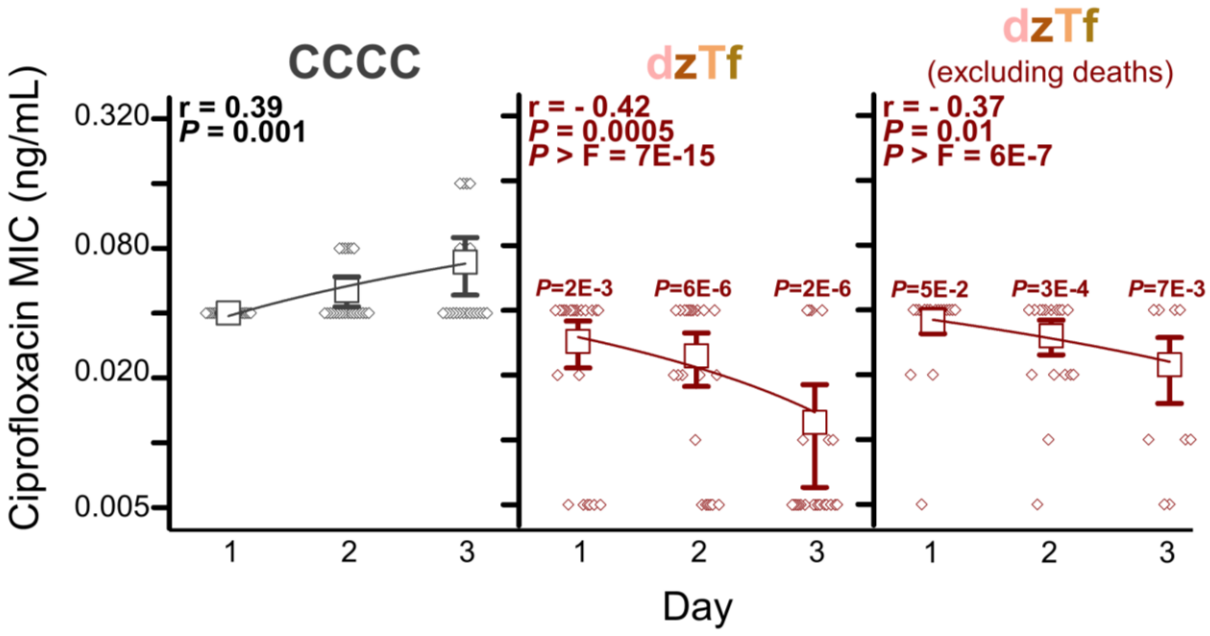
Supplementary Figure 6.16 Grubb's test removal of outlier

Ciprofloxacin MICs of the control strain run during the same experimental run as strains dfp, zwf, topA, frr, msTf and dzTf in Figure 4. During this experimental run, one replicate of the control strain grew significantly higher than the other replicates, and sequencing of this replicate revealed a mutation resistance (see Table S6). This replicate was removed using Grubbs' test for outliers on the left and was kept in calculations on the right. The left version of the linear fit was used for all F-tests presented in Figure 4 for strains dfp, zwf, topA, frr, msTf and dzTf.



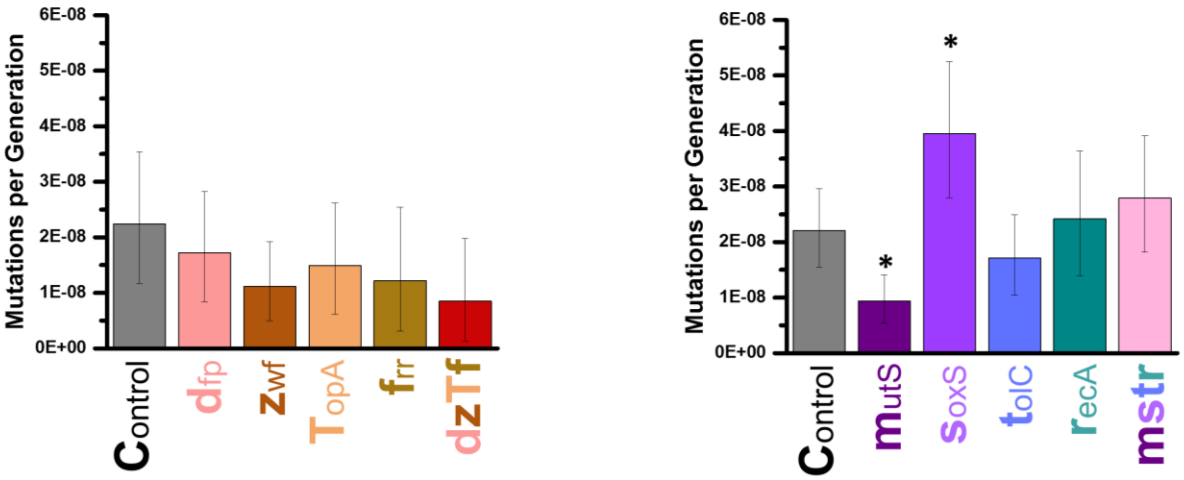
Supplementary Figure 6.17 Exclusion of replicates failing to grow in plain LB

Ciprofloxacin MICs from Figure 4 with slight modifications in their calculations. In Figure 4, replicates that failed to grow in absence of ciprofloxacin (i.e. replicate "deaths") were treated as having a MIC of 0.005 µg/mL (the lowest concentration tested). In this figure, such replicates were instead removed from the dataset. This affected 3, 6, and 15 replicates on days 1, 2, and 3 respectively for strain dzTf and zero, 1, and 3 replicates on days 1, 2, and 3 respectively for strain mstr. The probability of each linear fit being different than that of the control strain during the same experimental run was calculated using F-tests, and the resulting significance is presented as *P* > F on each graph. *P* values above each data average are in relation to the control from the same day of the experiment, using a two-tailed type II t-test.



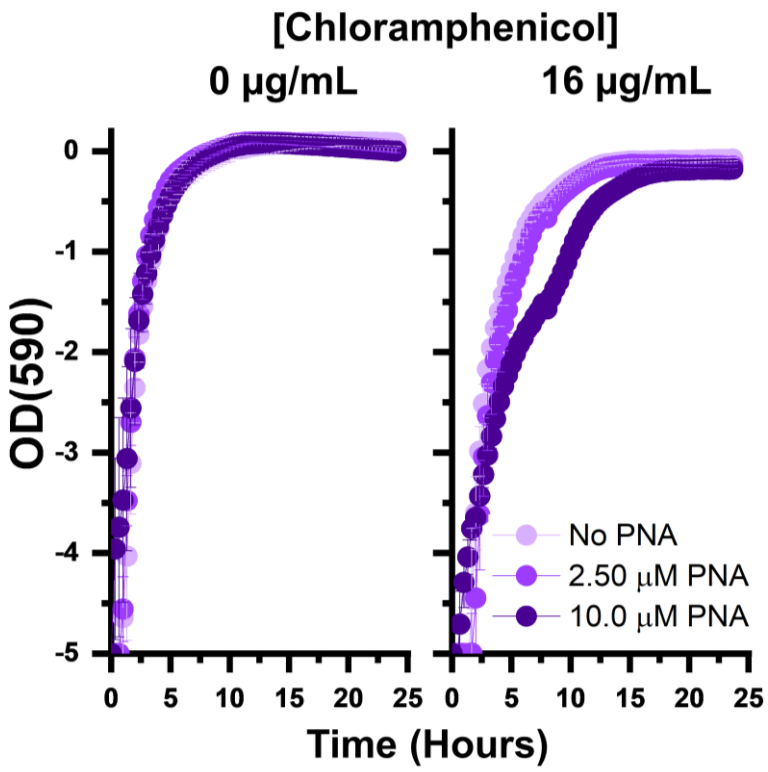
Supplementary Figure 6.18 Repeat experiment to test dzTf MIC

Average MICs of strain dzTf compared to the control in a separate experiment to confirm the results presented in Figure 4. Replicates of dzTf that failed to grow under the absence of ciprofloxacin exposure (i.e. replicate "deaths") were treated as having a MIC of 0.005 µg/mL (the lowest concentration tested) in the middle plot or were excluded from analysis in the right plot. This affected five, five, and 13 replicates on days one, two, and three of the experiment respectively. Statistical significance of Pearson Correlation Coefficients is listed underneath their corresponding fits. The probability that the Pearson Correlation Coefficient is statistically different from the control is listed as *P* > F on the two graphs of strain dzTf. *P* values above each data average are in relation to the control from the same day of the experiment, using a two-tailed type II t-test.



Supplementary Figure 6.19 Mutation rates of CHAOS strains

Mutation rates of strains as determined by mutation fluctuation assays outlined by Luria and Delbruck. Values indicate the number of mutations per generation that arise during one day of CRISPR perturbation. All comparisons were made in relation to the control strain from the same experimental run. The FALCOR web-tool was used to calculate mutation rates and error. Error bars indicate 95% confidence interval of at least 30 biological replicates. Asterisks indicate statistically different mutation rates relative to the control strain ($P < 0.05$) calculated using a two-tailed type II t-test



Supplementary Figure 6.20 CRE *E. coli* growth curves

Growth curves of the clinically isolated, multi-drug resistant CRE *E. coli* during exposure to various concentrations of PNA targeting a nonsense RNA sequence not present in the isolate. On the left is shown growth in the absence of chloramphenicol, and on the right is shown growth in the presence of 16 µg/mL chloramphenicol. A slight growth deficit was observed from the higher concentration of PNA only during chloramphenicol exposure. However, all samples were able to reach comparable densities by the end of 24 hours of growth, indicating that CFU values reported in Figure 6c were minimally impacted by inherently higher PNA concentrations, if at all. Error bars represent standard deviations of three biological triplicates.

Chapter 7: Conditional Killing: Engineering Fitness-Neutral Gene Perturbations to Resensitize Multidrug Resistant Bacteria to Treatment

Otoupal, P. B., Erickson, K. E., Eller, K. A., & Chatterjee, A. *In Preparation.*

7.1 Abstract

New strategies to combat multidrug-resistant (MDR) bacteria are desperately needed in order to prevent a post-antibiotic apocalypse. One approach is to develop co-therapies that enhance the efficacy of antibiotics. Here we systematically explore the rational engineering of gene expression to potentiate antibiotic therapy, without imposing a fitness cost when administered alone. We examined the synergy of 270 knockout-drug combinations in Escherichia coli, identifying ~100 promising interactions that improve antibiotic efficacy. This antibiotic synergy was replicated using CRISPR interference to knockdown expression of specific genes, which were subsequently multiplexed to enhance synergy and reduce the likelihood of mutational escape. We show these results are translatable to *Salmonella enterica*, where CRISPR interference potentiates antibiotic killing inside HeLa epithelial cells. Finally, we designed peptide nucleic acids (PNA) to knockdown gene expression in four MDR, clinically-isolated bacteria. These PNAs potentiate antibiotic trimethoprim and tetracycline activity, re-enabling these antibiotics to treat MDR bacteria. Our results highlight a promising avenue for the development of creative gene-specific treatment strategies that could potentially extend the utility of current antibiotic drugs.

7.2 Introduction

Antibiotic resistance is one of the foremost problems facing humanity. Estimates for the yearly cost imposed by antibiotic resistance reaches as high as $55 billion in the United States[14] and €1.5 across Europe[300]. Both the World Economic Forum[301] and the World Health Organization[287] warn of calamitous economic and health outcomes if current trends continue unabated; while approximately one million people die from such infectious yearly, deaths attributable to antibiotic resistant infections are estimated to skyrocket to 10 million annually by 2050[302]. This problem will only be exacerbated as multidrug-resistant (MDR) bacteria continue to emerge, necessitating the pursuit of alternative antimicrobial strategies.

Current research into antibiotics has focused largely on developing new drugs that exhibit bactericidal activity through novel mechanisms. This has led to what is often referred to as the antibiotic “arms race”, in which new drugs are continuously developed as bacteria continue to evolve new resistance[116,288,303]. It is only a matter of time before bacteria evolve resistance to these treatments, trapping us in a cycle of futility.

Here we introduce an alternative approach to combating antibiotic resistance with the potential to break this cycle. Rather than developing novel drugs utilizing alternative mechanisms for killing bacteria, we focus on devising therapies that function to potentiate the efficacy of current antibiotic treatment. Crucially, we design such therapies to have no direct impact on bacterial fitness when administered independently. Therefore, these treatments should impose no selective pressure for bacteria to evolve escape strategies that negate the treatments impact. However, in the presence of another antibiotic, they

have the capacity to act as a co-therapy and potentiate its lethality. The idea of potentiating antibiotic treatment is well-established in literature, and has been applied through such strategies as inhibiting β-lactamase to increase susceptibility to β-lactam antibiotics[304]. The drawback that such therapies have is that directly targeting the mechanism of resistance imposes a selective pressure that encourages further evolution of the resistance protein. An alternative approach is to supply multiple antibiotics that synergize with one another in a co-therapy approach. This has been extensively characterized between common antibiotics[305–309], but it has been demonstrated that this approach can accelerate the evolution of multidrug resistance[310,311]. To our knowledge, the design of synergistic therapies that exhibit no fitness impact when administered independently has never been considered.

The design of such therapies have been made possible with recent advances in synthetic biology and the corresponding development of novel tools with sequence-specific targeting capabilities. This includes genome and transcriptome editing based on clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonucleases (such as Cas9), which has been employed as an antimicrobial to selectively degrade particular genetic elements[80,296]. CRISPR has also been employed to interfere with expression of particular genes (CRISPRi) and has been explored for antibacterial applications[19]. Another tool for sequence-specific gene targeting are peptide nucleic acids (PNAs), single stranded DNA mimics that bind tightly to the corresponding antisense mRNAs[87]. PNAs have demonstrated effective antimicrobial activity when conjugated to a cell-penetrating-peptide (CPP) motifs and targeted to knockdown gene

expression[93,96,97,102]. Either of these approaches could be used to reduce expression of particular genes to enhance antibiotic killing.

To design fitness-neutral potentiating therapies based on these technologies, we must first understand what genes to target. To this end, we began with a systematic exploration of gene-drug synergy in the well characterized bacteria *Escherichia coli*, for which there exists a collection of every viable gene knockout[138]. Successful growth in the absence of these genes suggests that their loss provides minimal fitness impact, fulfilling our first criteria of designing therapies that themselves are harmless. Their existence also enables rapid screening of how analogous gene knockdowns will impact antibiotic efficacy. We focused on thirty stress-response gene knockouts that we identified to be differentially expressed during antibiotic exposure in our previous works[16,17]. We systematically characterized how each of the strains responded to growth in sub-minimum inhibitory concentrations (MICs) of nine commonly used antibiotics representing a broad spectrum of functional classes. From this we identified significant synergistic interactions and constructed CRISPR interference constructs to replicate their results individually and in a multiplexed fashion. These constructs were also applied to *Salmonella* infections of HeLa cells to demonstrate their therapeutic potential in amplifying antibiotic efficacy. Finally, we utilize PNAs to knockdown these genes in five clinically-isolated, MDR strains of bacteria. We show that this approach successful resensitizes the bacteria to antibiotic treatment without imposing its own fitness cost. Together, these results demonstrate a new platform for discovering and designing synergistic therapies for enhanced antimicrobial treatment, with reduced potential for evolutionary escape.

7.3 Results

7.3.1 Selecting gene targets and antibiotic concentrations

We selected thirty genes to evaluate as potential targets for antibiotic combination therapy (Table 1). These genes were chosen due to existing evidence of their association with stress response and/or adaptation processes, which are promising candidates for potentiating antibiotic activity[312]. We previously quantified the behavior of certain SOS response (*recA*, *polB*, *dinB*, *dam*), general stress response (*rpoS*, *hfq*, *cyoA*, *mutS*), and *mar* regulon (*marA*, *rob*, *soxS*, *acrA*, *tolC*) genes during adaptation[16]. Several genes were also found to impact adaptive resistance in a transcriptome-level analysis of adapted versus unadapted strains (*fiu*, *tar*, *wzc*, *yjjZ* were differentially expressed while *ybjG*, *ydhY*, *ydiV*, and *yehS* were differentially variable)[17]. Finally, we selected transcriptional regulators that control expression of these and other genes (*bglG*, *crp*, *csgD*, *flhC*, *flhD*, *fnr*, *fur*, *gadX*, and *phoP*). The selected genes represent diverse functionalities, including transport (*acrA*, *tolC*, *fiu*), mutagenesis (*mutS*, *dam*, *polB*, *dinB*), motility (*tar*, *flhC*, *flhD*, *ydiV*), general global regulation (*rpoS*, *marA*, *fnr*, *fur*, *gadX*, and others), and include a few targets with unknown function (*yjjZ*, *yehS*). Knockouts of gene were obtained from the Keio collection. As *E. coli* is still viable upon deletion of these genes, it is likely that future sequence-specific gene therapies that knockdown expression of these genes would impose minimal inherent fitness cost.

Table 7.1 The 30 nonessential genes of *E. coli* investigated in this study

Gene	Description	Function
polB	DNA polymerase II	**DNA processes**
recA	DNA strand exchange & recombination protein	
dam	DNA adenine methyltransferase	
dinB	DNA polymerase IV	
mutS	Methyl-directed mismatch repair	
wzc	Colanic acid biosynthesis protein	**Metabolism**
tar	Methyl accepting chemotaxis protein	**Motility**
flhC	FlhC-FlhD transcriptional regulator of flagellum biogenesis	
flhD	FlhC-FlhD transcriptional regulator of flagellum biogenesis	
ydiV	Anti-FlhDC factor	
bglG	Uptake and utilization of β-glucosides	**Regulation**
crp	cAMP receptor protein, regulates energy metabolism	
csgD	Regulates curlin genes, important for biofilm formation	
fur	Ferric uptake regulator	
gadX	Controls transcription of pH-inducible genes	
hfq	RNA-binding protein	
phoP	Two component regulatory system phoQ/phoP	
rob	Regulator induced by dipyridyl, bile salts, or decanoate	
rpoS	RNA polymerase, sigma S	
fnr	Regulator, mediates aerobic to anaerobic transition	
marA	Multiple antibiotic resistance regulator	**Redox**
soxS	Regulation of superoxide response regulon	
ydhY	Putative oxidoreductase system protein	
ybjG	Putative bacitracin resistance protein	
cyoA	Cytochrome bo terminal oxidase subunit II	**Transport**
tolC	AcrAB-TolC multidrug efflux pump - membrane fusion protein	
acrA	AcrAB-TolC multidrug efflux pump - membrane fusion protein	
fiu	Outer membrane receptor for iron transport	
yjjZ	Unknown	**Unknown**
yehS	Unknown	

We chose to test these knockouts growth in a set of nine antibiotics representing a diverse set of common antimicrobial therapies (Table 2). We initially performed growth assays to determine a suitable concentration for each antibiotic agent, in order to select a suitable dose that is less than the minimum-inhibitory concentration (MIC). We grew

wild type *E. coli* BW25113 in two-fold increments of each drug (see Methods) while quantifying growth inhibition per the ratio of the maximum optical density achieved during 16 hours with and without antibiotic treatment. Antibiotic concentrations resulting in a 10-50% fold reduction in growth were identified and used for all subsequent testing[305].

Table 7.2 The nine antibiotics investigated in this study and the working concentrations used

Antibiotic	Abbreviation	Dose (μg/mL)	Target
Ampicillin	Amp	2	Cell wall
Ceftriaxone	Ceft	2	Cell wall
Tetracycline	Tet	0.25	Protein synthesis (30S)
Erythromycin	Erm	50	Protein synthesis (50S)
Chloramphenicol	Cm	0.4	Protein synthesis (50S)
Puromycin	Pur	50	Protein synthesis (RNA)
Ciprofloxacin	Cip	0.008	DNA replication
Sulfadimidine	Sulf	0.5	DNA/RNA synthesis
Trimethoprim	Trim	0.125	DNA/RNA synthesis

7.3.2 Gene knockout synergy with antibiotic treatment

We characterized BW25113 growth in 270 combinations of 30 Keio knockouts with these nine antibiotics. A representative demonstration of how synergy was calculated is presented in Fig. 1A. In this example, BW25113 experienced no substantial fitness impact when grown in the presence of ampicillin (Wx = 1.04 ± 0.05) or in the absence of *acrA* (Wy = 0.93 ± 0.04). However, introducing this genetic change during exposure to ampicillin resulted in a dramatic drop in cell fitness (Wxy = 0.07 ± 0.01), showing clear potentiation of ampicillin efficacy by removing *acrA* (Synergy = Wx*Wy – Wxy = 0.89 ± 0.06). This process was repeated for all 269 remaining gene-drug pairs (Fig. 1B). Interactions were classified as synergistic if the lower bound of the calculated synergy was positive upon subtracting the standard deviation, or antagonistic if the upper bound was negative upon adding standard deviation. Other interactions were deemed additive.

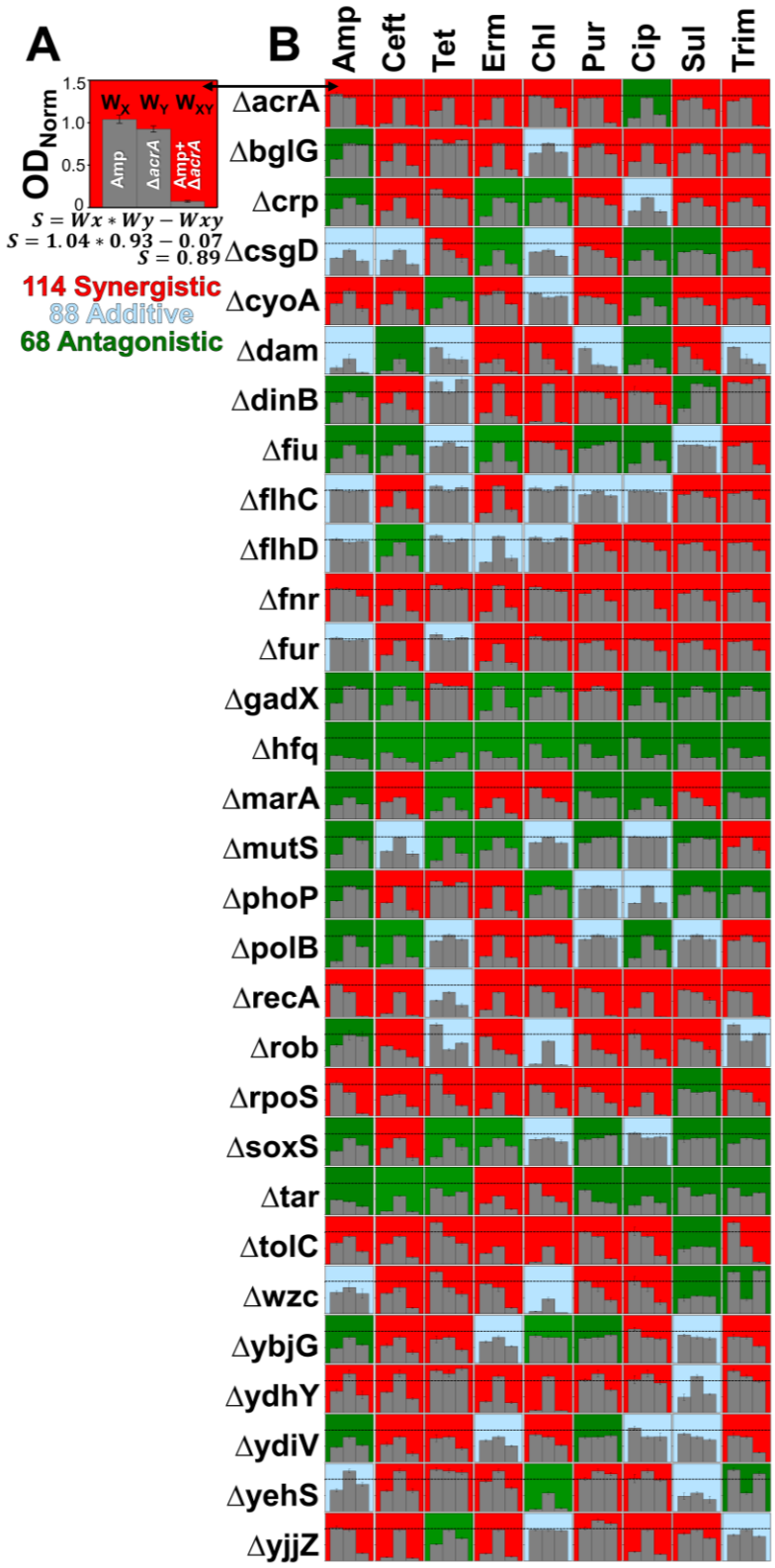


Figure 7.1 Fitness values used for calculating gene-drug synergy

(**A**) An example of how synergy values were calculated. The maximum growth of wildtype BW25113 during 16 hours was used to normalize growth of BW25113 in ampicillin, BW25113 with *acrA* deleted, and both simultaneously. These normalized growth values were used as a proxy for strain fitness (W) in the presence of antibiotic (Wx), gene knockout (Wy), or both simultaneously (Wxy). Synergy was calculated as Wx * Wy – Wxy, with positive values indicating synergy and negative values indicating antagonism. (**B**) This process was performed for all 270 gene-drug combinations. Interactions that proved significantly synergistic (or antagonistic) are color-coded red (or green). Non-significant interactions are classed as additive (blue). Scale on all bar graphs range from 1.5 to 0.0. A dashed line extends across all graphs indicating neutral fitness of 1.0. Error bars represent standard deviation of four biological replicates.

A few genes exhibited consistent synergy profiles across all antibiotics tested. Deletion of *acrA, bglG, fnr, fur, recA, rpoS, tolC,* and *ydhY* each improved antibiotic efficacy in at least seven conditions, suggesting that these genes are promising targets for improving general antibiotic efficacy. These gene knockouts was also relatively strong, many of which exhibited an average synergy greater than 0.20 (including *acrA*, *fnr*, *recA*, *rpoS*, *tolC* and *ydhY*). Our previous investigation of the impact of *acrA* deletion on adaptation to ampicillin and tetracycline showed overall decreases in fitness over the 3-day adaptation period[16], which is in line with expectation considering the deleterious synergy observed here. Also supporting these results are previous findings that knockouts of *tolC*, *acrA*, and *recA* have particularly high sensitivity to a variety of antibiotics[135]. Conversely, a few gene knockouts resulted in frequent antibiotic antagonism. Removing *gadX, hfq,* and *tar* each diminished antibiotic efficacy in at least seven conditions. These results for *hfq* are particularly surprising, as a correlation between Δ*hfq* and abolished induction of the RpoS regulon has been previously established[137].

The goal of this systematic investigation was two-fold, the first of which was to identify strong gene-drug synergistic interactions. The second goal was to ensure that these knockouts imposed minimal fitness costs on cell growth, to protect against the possibility of natural selection working against sequence-specific therapies designed to target these genes. To this end, we considered the average fitness cost of each of these genetic deletions in the absence of antibiotic treatment. We considered treatments that reduced fitness Wy below 0.70 to constitute a meaningful growth burden on the cell, and were therefore avoided in future experiments. This included Δ*dam*, Δ*rob*, Δ*hfq*, and Δ*tar*,

the latter two of which also demonstrated consistent antibiotic antagonism. We made an exception for Δ*marA* (Wy = 0.68), as this was one of the few strains that demonstrated significant synergy with ceftriaxone. Furthermore, previous studies of *marA* specific therapies have demonstrated that disruption of this gene's expression primarily extends lag times, while having minimal impact on growth rates in the absence of environmental stress[19].

We sought to understand if disruption of particular cellular processes was related to the degree of synergy observed by grouping synergies by pathway categories (Fig. 2). We also listed the knockouts that provided the three highest and lowest levels of synergy with each antibiotic. Notably, the one knockout directly affecting metabolism, Δ*wzc*, represented one of the top three synergistic knockouts in ceftriaxone, tetracycline, erythromycin, and ciprofloxacin, but was also one of the three most antagonistic knockouts in sulfadimidine and trimethoprim. Whole genome RNA-sequencing showed that *wzc*, a colanic acid biosynthesis gene, was overexpressed during ampicillin exposure[17], although no significant synergy was observed between Δ*wzc* and ampicillin in this experiment. And while the classes of antibiotics in which synergy was observed were diverse, clear antagonism emerged in the antibiotics related two DNA/RNA synthesis.

Again, these results highlight that the TolC-AcrA efflux pump knockouts demonstrated some of the highest levels of gene-drug synergy. Knockouts of at least one of these genes was always one of the three most synergistic genetic changes for all of the antibiotics tested, with the exception of the 50S targeting antibiotics erythromycin and chloramphenicol. However, even in these antibiotics, both knockouts resulted in

significant synergy. The remaining standout knockouts include Δ*recA,* Δ*dam*, and Δ*rpoS*, which demonstrated high synergy with four, two, and two antibiotics respectively. There was no clear relationship between the cellular processes these genes are involved in and the antibiotics' modes of action.

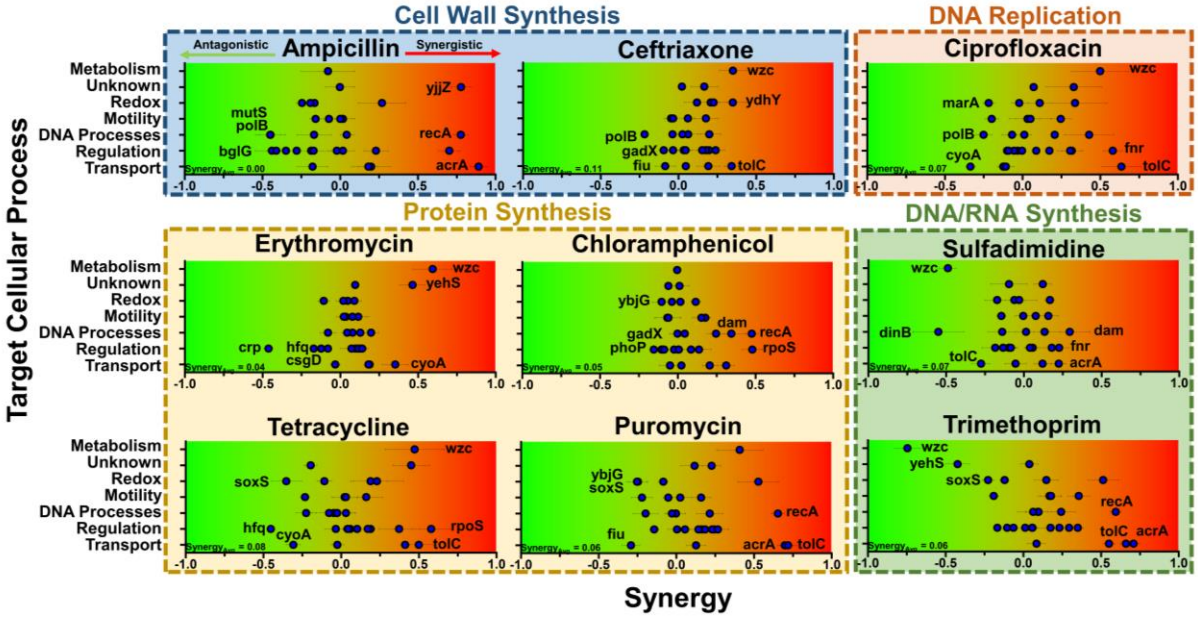


Figure 7.2 Degree of synergy between gene knockouts and antibiotic treatments

The synergy values of gene knockouts with each antibiotic are plotted by the treatment tested. Gene knockouts are separated into their specific cellular processes on the y-axis, with corresponding synergy plotted on the x-axis. Antibiotics are further grouped based on the mechanism of action, such as targeting cell wall synthesis. The top three synergistic interactions and top three antagonistic interactions are specifically labeled in each graph. In the bottom left of each graph is listed the average synergy of all thirty gene knockouts with the specific antibiotic. Error bars represent standard deviation of four biological replicates propagated from fitness values.

We next sought to ascertain potential mechanisms contributing to the degree of synergy induced by gene knockouts. One hypothesize is that the degree of synergy is influenced by the epistatic interactions the targeted gene is involved in, and the cascading consequences that knocking out said gene has throughout the cell. We utilized information from the STRING database of all known protein-protein interactions of each

of the thirty genes investigated here to find correlations between this information and the degree of synergy observed[280]. Protein interaction networks were constructed for each of the genes, including information on the total number of protein partners (nodes), the number of interactions between each protein in the network (edges), the average interaction of each protein in the network (node degree), and a measure of how connected (i.e. “tight”) nodes in the network are (clustering coefficient). Significant correlations were observed between a gene’s level of synergy and all of the aforementioned characteristics when analyzing the 270 gene-drug synergy scores (Supplementary Figure S1). Greater synergy was observed for genes involved in more genetic interactions (P = 9E-4), for networks with more internetwork connections (P = 2E-4), for more average interactions for each protein in the network (P = 0.012), and for networks with less overall tightness (P = 0.045). These results strongly suggest a role of the targeted gene’s genetic partners in influencing the degree of antibiotic potentiation possible.

Collectively, these results point to promising gene targets whose expression could be manipulated to enhance antibiotic efficacy. However, engineering the DNA of bacterial infections in order to knockout particular genes is unlikely to be a tangible therapeutic strategy. As such, we next turned our attention to translating these knockout results into more clinically relevant approaches.

7.3.3 Introducing gene-drug synergy using CRISPRi

If knocking out these genes resulted in significantly amplifying antibiotic potency, then it stands to reason that lowering their expression without completely removing them from the genome might engender similar results. Furthermore, this approach is a more

practical application of gene-drug synergy than performing genome editing to delete specific genes. To this end, we turned to CRISPRi to knockdown expression of genes showing significant antibiotic synergy (Fig. 3A). By utilizing catalytically dead Cas9 (dCas9) to reduce mRNA production, we sought to develop constructs that replicated gene knockout synergy in a therapeutically viable manner.

We specifically focused on the six genes showing the greatest degree of synergy with each of the antibiotics tested. We utilized a dual-plasmid system based on the original CRISPRi system to deliver gene knockdown constructs to BW25113[214]. One plasmid encoded expression of dCas9 under the aTc inducible promoter, while the other encoded a unique single guide RNA (sgRNA) target to repress expression. These targets were designed to target either with the first ~50 nucleotides of the gene's open reading frame, or around the promoter sequence to interfere with RNA polymerase binding. As these two plasmids rely on ampicillin and chloramphenicol selection markers, we excluded exploration of these antibiotics going forward. Additionally, due to the general low degree of synergy demonstrated by gene knockouts with sulfadimidine, this antibiotic was not included.

The gene-antibiotic synergy experiments were again repeated, with CRISPRi employed in the place of gene knockouts. A control strain in which the sgRNA target was aimed at the non-existent red fluorescent protein (*rfp*) gene was used in lieu of wildtype BW25113. Fitness was measured for this strain exposed to each antibiotic, each individual perturbation strain in the absence of antibiotic, and the perturbation exposed to antibiotic. All fitness values were compared to the fitness of the control strain the absence of any antibiotic (Fig. 3B). Notably, every CRISPRi strain exhibited a slightly higher fitness

than the control strain; the median Wy was 1.16, indicating that there was an inherent overestimation of the CRISPRi strains. While this strongly suggests that CIRSPRi imposed no inherent fitness cost when targeted towards these non-essential genes, it does reduce the apparent impact of the gene perturbations on antibiotic efficacy. For instance, the fitness Wx of the control in the presence of puromycin was 0.86 ± 0.07, but the fitness Wy of the CRISPRi of *bglG* was 1.25 ± 0.33. Although the fitness of *bglG* knockdown in the presence of puromycin was 0.85 ± 0.08, the calculated degree of synergy between these treatments was 0.22 ± 0.30, a statistically significant difference from the null-hypothesis of an additive interaction when 22 replicates were considered.

Taken this into account, most of the CRISPR perturbations resulted in statistically significant synergy. All of the tested perturbations improved efficacy of ciprofloxacin, trimethoprim, and erythromycin. Four of the six perturbations synergized with ceftriaxone (inhibition of *soxS, tolC, phoP,* and *marA*), and three of the six perturbations synergized with puromycin (*acrA, tolC,* and *bglG*) or tetracycline (*acrA*, *tolC*, and *csgD*). The only combination that resulted in clear antagonism was *rpoS*-i during tetracycline exposure. A few CRISPR perturbations did stand out from the rest in the very clear antibiotic synergy they induced. Most notable is the degree of synergy induced by inhibitions of the *tolC-acrA* efflux pump in erythromycin (*acrA*-i = 0.41 ± 0.12, *tolC*-i = 0.39 ± 0.09), puromycin (*acrA*-i = 0.72 ± 0.30, *tolC*-i = 0.67 ± 0.25), tetracycline (*acrA*-i = 0.69 ± 0.40, *tolC*-i = 0.61 ± 0.21), and trimethoprim (*acrA*-i = 0.78 ± 0.14, *tolC*-i = 0.72 ± 0.13). Inhibitions of *recA* and *fnr* additionally showed significant improvements in ciprofloxacin efficacy (synergy = 0.90 ± 0.14 and 1.06 ± 0.15 respectively).

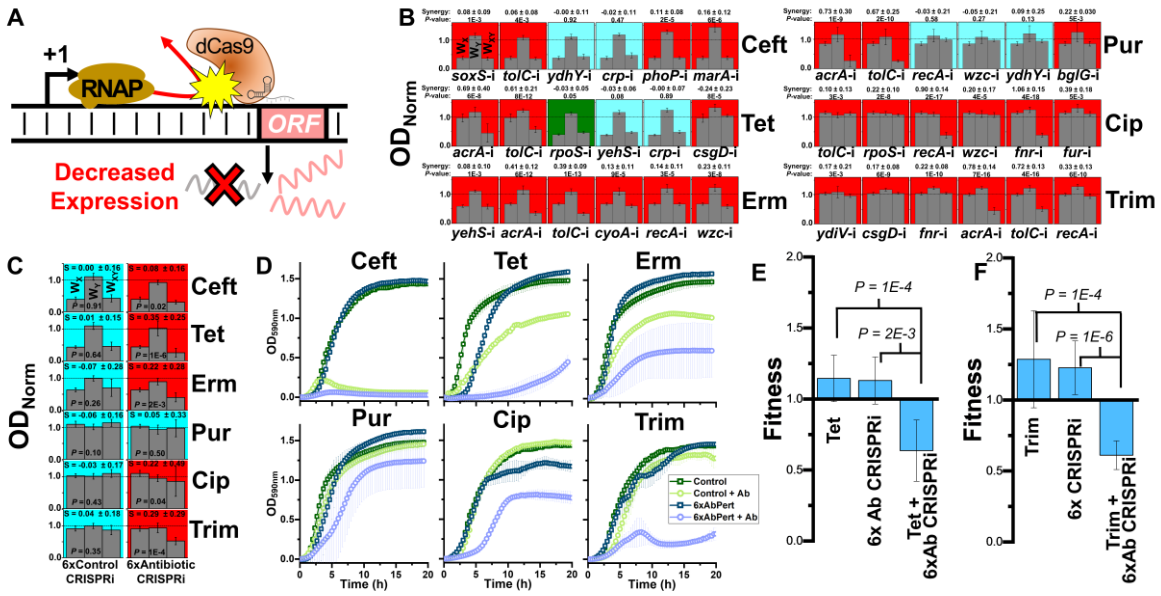


Figure 7.3 Applying CRISPRi to potentiate antibiotic treatment

(**A**) dCas9 is targeted to promoter or open reading frame elements of specific genes, preventing RNA polymerase from transcribing DNA into mRNA. Constructs were created to block transcription of six genes for which deletion resulted in significant synergy with a specific antibiotic. (**B**) Each of these CRISPRi strains were tested for their synergy with antibiotic treatment. Significant synergy or antagonism is indicated by a red or green background respectively, with additive interactions shown in blue. Synergy values are listed above each graph with significance. Error bars represent standard deviation of 22 biological replicates. (**C**) The six gene perturbations for each antibiotic were multiplexed into one strain, and synergy was again screened for (right column). A control strain with six nonsense *rfp* perturbations was also created to show that harboring multiple targets did not influence these results (left column). Error bars represent standard deviation of 22 biological replicates. Synergy values are listed above each graph with significance. (**D**) Growth curves of these multiplexed CRISPRi strains in the presence of each antibiotic. Error bars represent standard deviation of three biological replicates. A more thorough fitness assay using competition was applied to more precisely estimate the fitness impacts of multiplexed perturbations for (**E**) trimethoprim and (**F**) tetracycline. Competition was performed for these strains against a fluorescent control strain harboring one nonsense CRISPRi perturbation in either the presence or absence of antibiotic treatment. A control competition of the 6x *rfp* perturbation strain against the fluorescent control was also performed in the presence of antibiotic. Error bars represent standard deviation of eight biological replicate.

While CRISPRi appeared to largely successfully replicate gene knockout synergy with antibiotic treatment, the degree of synergy appears to be relatively low in comparison

to the synergy observed by gene knockout. The average synergy of these perturbations was lower than that of their corresponding knockouts in ceftriaxone (CRISPRi = 0.06 ± 0.24, Δ = 0.26 ± 0.15), tetracycline (CRISPRi = 0.25 ± 0.33, Δ = 0.42 ± 0.24), erythromycin (CRISPRi = 0.23 ± 0.27, Δ = 0.32 ± 0.17), puromycin (CRISPRi = 0.27 ± 0.62, Δ = 0.54 ± 0.22), and trimethoprim (CRISPRi = 0.40 ± 0.33, Δ = 0.49 ± 0.12). The only exception to this result was the perturbations in ciprofloxacin (CRISPRi = 0.48 ± 0.36, Δ = 0.40 ± 0.24). It stands to reason that knocking down gene expression would create less of an impact than complete removal of said gene, so these lower levels of synergy are to be expected.

To have a better understanding of these perturbations on antibiotic efficacy, we performed growth curve analysis of each CRISPRi strain with their corresponding antibiotics (Supplementary Figures S2-S7). In most cases, synergy between CRISPRi and antibiotics is made apparent in the early stages (5-10 hours) of growth. These results highlight that CRISPRi can effectively potentate antibiotic treatment.

7.3.4 Multiplexing CRISPRi exacerbates antibiotic synergy

A great benefit of the CRISPRi approach is the relative ease in which individual perturbations can be combined together in a single cell by including multiple sgRNAs. Furthermore, we have previously shown that multiplexing perturbations tends to exacerbate detrimental fitness impacts by inducing negative epistatic interactions between the perturbed genes[20]. This suggests that multiplexing synergistic CRISPRi perturbations could exacerbate the potentiation of antibiotic efficacy. We took advantage of this by combining the six aforementioned perturbations designed for each antibiotic

into one construct, and tested the resulting impacts on *E. coli* growth during antibiotic exposure.

We first ensured that expanding the number of perturbations did not have an inherent impact on growth by testing the growth of a control strain harboring six *rfp* perturbations (Fig. 3C). This strain exhibited no significant shift in basal fitness (Wy = 1.03 ± 0.10 on average), nor did it show antagonism or synergy with any antibiotic. However, all of the multiplexed CRISPRi perturbation strains designed for inducing synergy showed significant potentiation of antibiotic efficacy, with the exception of puromycin perturbations. Synergy was particularly pronounced with tetracycline (0.35 ± 0.25, P = 1E-6), erythromycin (0.22 ± 0.28, P = 2E-3), and trimethoprim (0.29 ± 0.29, P = 1E-4). The strong synergy observed by multiplexed tetracycline perturbations is particularly notable given the relative lack of synergy observed when perturbations were applied individually.

To further elucidate these multiplexed perturbations impacts on growth, we examined each strain's growth profile over 20 hours in both the presence and absence of antibiotic, and compared these profiles to the six perturbation control strain (Fig. 3D) All strains grew the same as the control in the absence of antibiotic exposure, with the exception of the six perturbation tetracycline strain which exhibited a slight increase in lag times. This strain was still able to reach similar growth levels as the control strain by roughly 10 hours of growth. In the presence of antibiotics, all antibiotic perturbation strains clearly demonstrated diminished growth capacity compared to the control strain, with the exception of puromycin perturbations. This supports the previous synergy calculations. The most notable shifts in growth profiles in antibiotic exposure arose from the

tetracycline and trimethoprim perturbations, suggesting these multiplexed perturbations induce the greatest potential antibiotic synergy. We utilized a more thorough experiment to analyze the fitness impacts of these perturbations using a competition assay against a control strain harboring one CRISPRi *rfp* perturbation and constitutively expressing mCherry (Fig. 3E,F). This showed no significant fitness reduction of these strains in the absence of antibiotic exposure, while still clearly potentiating antibiotic efficacy (Fitness = 0.64 ± 0.22, *P* = 1E-4 and Fitness = 0.61 ± 0.10, *P* = 1E-4 for tetracycline and trimethoprim multiplexed perturbations respectively). Collectively, these results demonstrate how multiplexed perturbations can further the potentiation of antibiotic efficacy.

7.3.5 CRISPRi potentiates antibiotic efficacy in infection models

A major benefit of the CRISPRi strategy for enacting sequence-specific gene therapies is the relative ease at which it is applied to a vast array of organisms. For instance, many of these CRISPRi constructs can be directly applied to a relative of *E. coli*, the bacteria *Salmonella enterica* serovar Typhimurium SL1344. Analysis of the SL1344 genome shows that six perturbations in particular have complete (*acrA*, *cyoA*, and *fnr*) or near-complete (*crp*, *rpoS*, and *tolC*) homology, and should therefore maintain efficacy even in this new organism (Fig. 4A). The three genes showing near-complete homology only exhibit single mismatches towards the 3’ end of the sequence, well outside the seed sequence of the sgRNA and therefore unlikely to abolish the ability of dCas9 to bind to these genes. Furthermore, SL1344 is a model organism for studying bacteria in intracellular infections due to the relative ease in which it infects human cell lines. To explore the potential for gene expression perturbations to potentiate antibiotic efficacy in

a therapeutic context, we created a new strain of SL1344 in harboring CRISPRi perturbations of these six genes. A strain of SL1344 harboring just the three perturbations with perfect homology, as well as a strain harboring six nonsense *rfp* perturbations were created as well.

We tested the synergy of the three and six perturbation strains for each of the six antibiotics (Fig. 4B). No detrimental impact on basal SL1344 fitness was observed in the absence of antibiotics for either of these strains (Wy = 1.14 ± 0.001 and Wy = 0.97 ± 0.07 for the three and six perturbations respectively). However, significant synergy was observed in a number of instances, strongly suggest that our CRISPRi system was successfully applied across organisms as a potentiators of antibiotic activity. Both strains showed significant synergy with ceftriaxone and tetracycline, the latter of which appeared to be particularly potentiated (synergy = 0.33 ± 0.24, *P* = 3E-6 and 0.19 ± 0.13, *P* = 9E-7 for three and six perturbations respectively). Very strong synergy was also observed between trimethoprim and the six perturbation strain (0.28 ± 0.13, *P* = 1E-9). Slight synergy was observed with erythromycin (0.13 ± 0.12, *P* = 3E-5) and puromycin (0.14 ± 0.15, *P* = 4E-4) for the three perturbation strain, but this synergy was lost upon combining the remaining three perturbations. Additionally, slight antagonism was actually observed between the six perturbations and ciprofloxacin (-0.09 ± 0.10, *P* = 3E-4), which could be due to the antagonism demonstrated by Δ*acrA* and Δ*cyoA* in BW25113. As these three antibiotics had very little or no potentiation of their activity, they were not explored further. Going forward, we focused our efforts on characterizing these perturbations' impacts on tetracycline and trimethoprim, as these antibiotics showed the highest level of synergy.

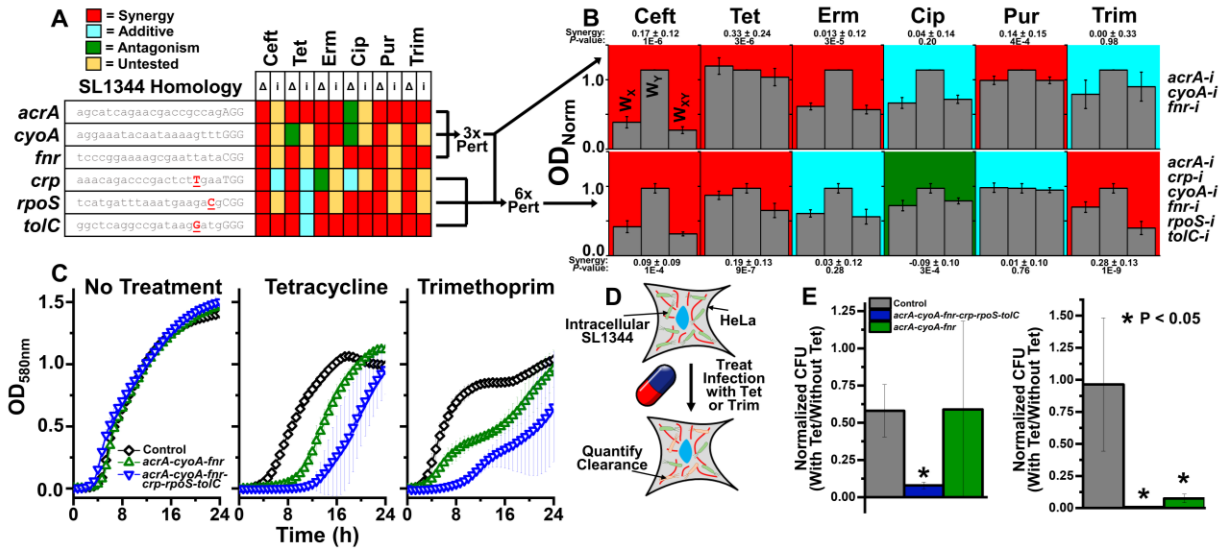


Figure 7.4 CRISPRi potentiation of antibiotic treatment of intracellular *Salmonella* infections

(**A**) The exact 20nt target sequences of six CRISPRi constructs are listed, with the native PAM sequence listed in capitals at the end of each sequence. Underlined red sequences indicate a mismatch in the sgRNA sequence with the native sequence of *Salmonella enterica* serovar Typhimurium SL1344. On the right is shown how these gene knockouts (Δ) or CRISPRi knockdowns (i) interacted with the corresponding antibiotic. (**B**) Two CRISPRi constructs targeting the genes with perfect homology (*acrA*, *cyoA*, and *fnr*, top row) or all six genes (bottom row) were created and screened for their ability to potentiate antibiotic treatment of SL1344. Significant synergy (or antagonism) is indicated with a red (or green) background, with blue representing additive interactions. Synergy values are listed above each graph with significance. Error bars represent standard deviation of 22 biological replicates. (**C**) Growth curves of CRISPRi SL1344 strains in the presence or absence of antibiotic treatment. Error bars represent standard deviation of five biological replicates. (**D**) CRISPRi treatments were applied to SL1344 during intracellular infection of HeLa cells to observe their ability to potentiate antibiotic treatment in a clinically relevant environmental setting. (**E**) Survival of CRISPRi SL1344 strains in intracellular HeLa infections after 18 hours of 0.5 µg/mL tetracycline or 0.5 µg/mL trimethoprim treatment, relative to survival of the same strain during infection with no antibiotic treatment. *P* values are given in relation to the control strain. Error bars represent standard deviation of three biological replicates and two technical duplicates.

The growth profiles of these perturbed SL1344 strains were compared to the control strain of SL1344 in the presence of no antibiotic, tetracycline, or trimethoprim (Fig. 4C). No impact on growth was observed in the absence of antibiotic exposure, indicating

these perturbations imposed no direct fitness cost. Clear detrimental impacts on growth were observed during tetracycline and trimethoprim, supporting the idea that these perturbations potentiated antibiotic treatment of SL13444. While our previous results showed a lack of synergy between the three perturbations and trimethoprim, this appears to be explained by the eventual ability of the strain to reach similar growth levels as the control strain.

These SL1344 strains were then used to infect HeLa epithelial cells to investigate the ability of perturbations to potentiate antibiotic clearance of intracellular infections. Infected HeLa was subjected to no antibiotic, 0.5 µg/mL tetracycline, or 0.5 µg/mL trimethoprim for 18 hours post infection. HeLa were then lysed, and colony-forming units of remaining SL1344 were determined. The surviving SL1344 in the presence of antibiotic were compared to the relative surviving salmonella in the absence of antibiotic (Fig. 4D). Significant reductions in the relative amount of viable SL1344 were observed in the presence of trimethoprim for both perturbation strains when compared to the control; cell survival of the six perturbation strain was reduced 99.2% (P = 0.03) and survival of the three perturbation strain was reduced 92.2% (P = 0.04) (Fig. 4E). The six perturbation strain also experienced significant reduction in survivability in tetracycline, where an 86.4% reduction in colony forming units was observed (P = 0.008). These results indicate that the targeted multiplexed CRISPRi constructs successful potentiated intracellular antibiotic treatment, supporting the therapeutic viability of fitness-neutral gene perturbation treatments.

7.3.6 PNA knockdown of gene expression resensitizes MDR clinical isolates to antibiotic treatment

To further explore the therapeutic potential of fitness neutral gene perturbations, we utilized an alternative gene expression knockdown approached based on PNA. The structure of PNA and DNA are very similar (Fig. 5A), and the ability of PNA to bind to single stranded DNA and RNA has been well established. When conjugated to cell-penetrating-peptides (CPPs), PNAs can readily cross bacterial membranes and enter the cell. We designed a set of 12nt long PNAs conjugated to a CPP motif $[KFF]_3K$ to target the start codon of genes showing strong antibiotic synergy in *E. coli* and *Salmonella*. When these PNAs enter the cell, they form tight bonds with complementary mRNA, preventing ribosome translation of these genes into proteins. The benefit of this approach is that it can be readily applied to a wide array of bacteria without requiring the cloning of plasmids to enact gene expression knockdown. Therefore, we can readily utilize PNAs to stimulate gene-drug synergy in various strains of clinically isolated MDR resistant bacteria.

We applied PNA knockdown of gene expression to four such strains that exhibit a wide range of resistances. This includes a two strains of MDR *E. coli*, one of which exhibits a carbapenem-resistant Enterobacteriaceae (CRE) phenotype, and two strains of *Klebsiella pnuemoniae* (KPN) producing either an extended spectrum β-lactamase (ESBL) or a New Delhi Metallo β-lactamase 1 (NDM-1). These strains are able to survive a wide range of antibiotic concentrations significantly above the resistance breakpoint levels established by the clinical & laboratory standards institute (Fig. 5B). This is significantly better growth than wild-type BW25113 growth in trimethoprim (Fig. S8).

We chose to focus specifically on enacting synergy with trimethoprim, as we achieved the greatest success in engineering synergy with this antibiotic in our CRISPRi approach. We first screened the genomes of these four strains for homology with each of the six trimethoprim gene perturbations tested with CRISPRi. Low homology was found for *YdiV*, so we chose to exclude testing of this gene. The remaining five genes showed significant homology with the four MDR strains. However, we chose to exclude testing *tolC* as well, as multiple off-targets were found. Additionally, we wanted to target as diverse a range of cellular pathways as possible, meaning that the significant overlap between the *tolC* and *acrA* PNA targets was undesirable. We designed PNA molecules inhibiting expression of these four genes by targeting them to overlap the start codon of each gene's ORF, as well as a PNA targeting a nonsense sequence not present in any of the genomes. Testing the growth impact of this nonsense PNA on growth of each of the MDR strains revealed that it had minimal impact on strain growth in the presence of trimethoprim, indicating that PNA or CPP molecules alone had no significant impact on MDR growth or antibiotic potentiation (Fig. S9).

We next examined the ability of the four targeted PNAs to synergize with trimethoprim treatment in each of the four MDR strains. While trimethoprim treatment alone demonstrated no effect on any MDR growth, two of the PNAs exhibited significant impact on MDR growth in the absence of trimethoprim: the *recA* targeting PNA reduced growth of all strains but NDM-1 KPN, while the *csgD* PNA reduced growth of ESBL KPN and CRE *E. coli*. While significant reduction in growth was observed in each of these cases in the presence of trimethoprim as well, their inherent fitness impact does not conform to our fitness-neutral perturbation design criteria.

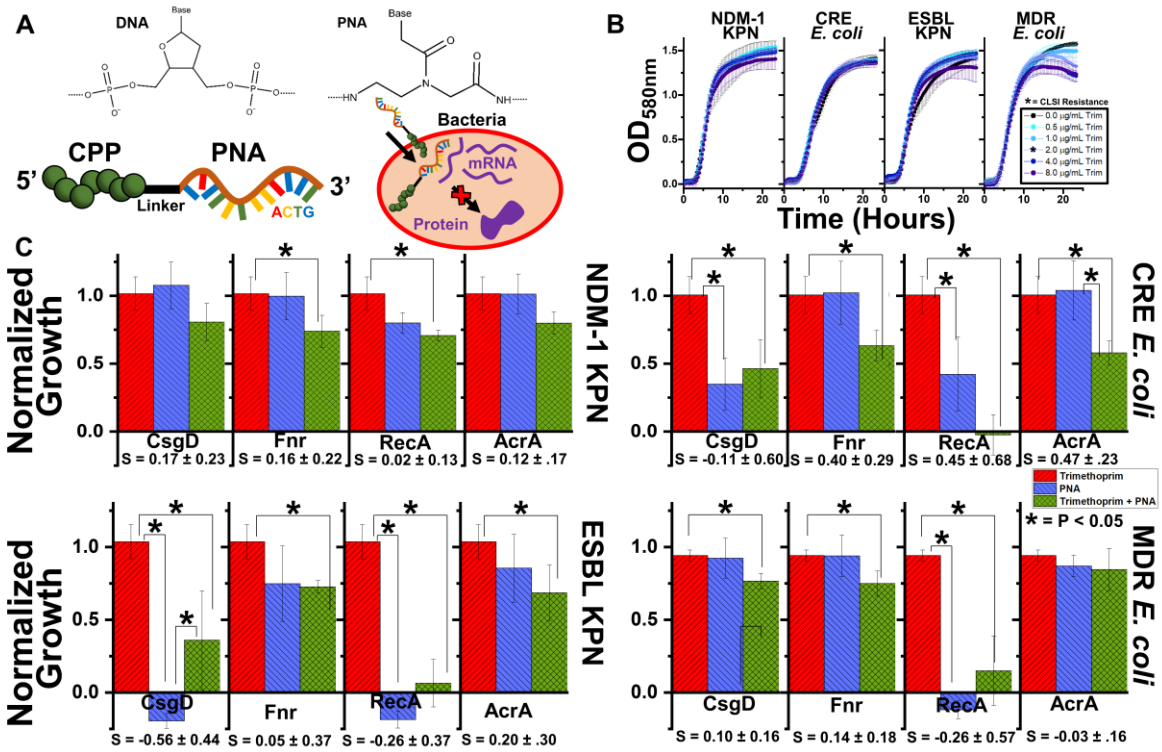


Figure 7.5 PNA gene knockdown treatment resensitizes MDR clinical isolates to antibiotics

(**A**) Chemical structures of DNA and PNA show how the negatively charged phosphate backbone of DNA is replaced with a neutrally charged peptide backbone in PNA. These PNAs are conjugated to CPP to enable penetration of bacterial membranes. Upon entry to the cell, PNAs complex with complementary mRNA to inhibit protein translation. (**B**) Resistance of MDR, clinically isolated bacteria to trimethoprim above CLSI cutoff levels of resistance, as demonstrated by growth curves unaffected by trimethoprim concentration. Error bars represent standard deviation of four biological replicates. (**C**) MDR bacteria growth after exposure to 2 µg/mL trimethoprim (red bars), 10 µM PNA (blue bars), or both simultaneously (green bars). Growth is normalized to growth in the absence of treatment. Synergy values of PNA with trimethoprim are given below each bar chart. Error bars represent standard deviation of biological triplicates.

There were a number of instances in which PNA treatment did cause significant potentiation of trimethoprim without basal fitness impacts. NDM-1 KPN proved the most difficult to treat, but trimethoprim treatment was still potentiated by the *fnr* and *recA* PNAs ($P = 0.048$ and $P = 0.01$ respectively). While both ESBL KPN and CRE *E. coli* experienced trimethoprim potentiation upon exposure to every PNA, only the *fnr* ($P = 0.01$ and $P =$

0.02 respectively) and *acrA* PNAs (*P* = 0.01 and *P* = 0.03 respectively) exhibited no significant impact on cell growth in the absence of trimethoprim. Both *csgD* and *fnr* PNA potentiated MDR *E. coli* trimethoprim treatment (*P* = 0.038 and *P* = 0.01 respectively) while exhibiting no impact in the absence of trimethoprim. Looking at the 48 PNA-trimethoprim interactions at large, the average degree of antibiotic potentiation was significantly synergistic (S = 0.086 ± 0.32, *P* = 0.035). Together, these results indicate that targeted perturbations of nonessential genes can significantly potentiate antibiotic treatment of clinically isolated MDR bacteria, even in instances where they exhibit no significant impact on cell growth independently.

7.4 Discussion

Here we have demonstrated that sequence-specific gene expression therapies targeting nonessential genes can be engineered to potentiate antibiotic therapy without directly affecting bacterial fitness. We outline a discovery pipeline for how similar gene-drug synergies can be identified and exploited, building upon knowledge of gene knockouts to engineer CRISPRi or PNA gene knockdowns. We demonstrate this approach works in both an intracellular infection context as well as upon clinically isolated MDR bacteria.

While previous studies have employed the Keio library to explore the impacts of gene knockouts on antibiotic efficacy[135,313,314], these studies have not considered or explored how identified gene-drug synergy can be therapeutically exploited to potentiate antibiotic efficacy, as we have done here. Furthermore, these studies have not taken into consideration the fitness impact of gene knockouts in the absence of stress. There is

ample reasoning to suspect that fitness neutral therapies are highly preferable to those that diminish fitness in the absence of antibiotics, as synergistic interactions between antibiotics has been correlated with accelerated evolution of antibiotic resistance[310,315]. Likewise, previous studies examining how single gene disruption alters antibiotic fitness have lacked direct comparison to wild type controls, meaning that deleterious interactions cannot be defined as additive, synergistic, or antagonistic[135,313]. A lack of a wild type performance data prevents other large-scale explorations of bacterial phenotype from being capable of typifying gene/drug interactions[316]. The work presented here, and similar efforts that use CRISPR-interference to perturb gene expression in the presence of stressful environments[19], are critical to explore the widespread potential that synergistic, antagonistic, or additive interactions may have.

The targeting of nonessential pathways to combat antibiotic resistance is a lamentably underexplored strategy in the literature, despite its potential for potentiating treatment without having an effect on pathogens on their own[317]. While essential genes interference therapies can be developed, the likelihood that their inherent deleterious effects on fitness will encourage the evolution of new resistance requires significant consideration. The potentiating CRISPRi and PNA knockdowns of nonessential gene expression provides a promising approach to enhance our ability to treat MDR bacteria in the clinic without selecting for further resistance. This strategy is supported by a similar approach in which interference of LexA activity was applied to reduce expression of the nonessential genes *dinB*, *polB*, and *umuD*, resulted in significant potentiation of long-term ciprofloxacin and rifampicin treatment[318].

Finally, the pool of nonessential genes that can be further explored for gene-drug synergies is vast and remains largely untapped. Here we explored only 30 of the reported 3985 nonessential genes of *E. coli*[138], many of which could hold promise as targets for fitness neutral potentiation of antibiotic therapy. Ongoing work in our lab has also suggested that multiplexing gene perturbations can restrict the evolvability of bacteria, suggesting that combinatorial knockdown approaches might further enhance the results presented here[20,319]. As a whole, we are highly optimistic about the potential of gene/drug combinations therapies to realize promising candidates with clinical relevance for combating antibiotic resistance.

7.5 Materials and Methods

7.5.1 Target gene selection

We selected thirty genes to evaluate as potential targets for combination therapy. Many of these genes were chosen due to existing evidence of their association with stress response and/or adaptation processes. We previously quantified the behavior of certain SOS response (recA, polB, dinB, dam), general stress response (rpoS, hfq, cyoA, mutS), and mar regulon (marA, rob, soxS, acrA, tolC) genes during adaptation[16]. Several of the genes selected here were also found to impact adaptive resistance in a transcriptome-level analysis of adapted versus unadapted strains (fiu, tar, wzc, yjjZ were differentially expressed while ybjG, ydhY, ydiV, and yehS were differentially variable)[17]. Finally, we looked upstream and selected transcriptional regulators that control the expression of these and other genes (bglG, crp, csgD, flhC, flhD, fnr, fur, gadX, and phoP). The selected genes represent diverse functionalities, including transport (acrA, tolC, fiu), mutagenesis (mutS, dam, polB, dinB), motility (tar, flhC, flhD, ydiV), general global regulation (rpoS, marA, fnr, fur, gadX, and others), and include a few targets with unknown function (yjjZ, yehS).

7.5.2 Bacterial strains and culture

All knockout strains used are from the Keio collection[138]. The parent strain (Escherichia coli BW25113) and individual gene knockouts were obtained from Yale's Coli Genetic Stock Center (http://cgsc.biology.yale.edu/index.php). *E. coli* NEB10β was used for cloning of all CRISPR plasmids used in this study. Experiments using plasmids were done by transforming these plasmids into BW25113. *Salmonella enterica* serovar

Typhimurium SL1344 with genome-integrated Green Fluorescent Protein (GFP) was used for harboring CRISPR plasmids for intracellular infections.

Clinical isolates of multidrug resistant bacteria were obtained from the lab of Dr. Nancy Madinger at the University of Colorado Anschutz campus. This includes carbapenem-resistant Enterobacteriacea (CRE) *E. coli*, another multidrug resistant *E. coli*, *Klebsiella pnuemoniae* harboring New Delhi metallo-β-lactamase 1 (NDM-1), *Klebsiella pnuemoniae* harboring extended-spectrum β-lactamase, and multidrug resistant *Salmonella enterica* serovar Typhimurium.

Unless otherwise noted, all experiments using these strains were performed in lysogeny broth (LB), M9 minimal media supplemented with 2.0 mM $MgSO_4$ and 0.1 mM $CaCl_2$ and 0.4% glucose, or cation-adjusted Mueller Hinton Broth (caMHB). Keio collection strains were grown in the presence of 40 µg/mL kanamycin. Strains harboring CRISPR plasmids were grown in the presence of 100 µg/mL ampicillin and 35 µg/mL chloramphenicol, and supplemented with 50 ng/mL aTc for induction of dCas9 when appropriate. All liquid cultures were grown at 37 °C with 225 rpm shaking, and all plating was performed at 37 °C.

7.5.3 Determining sub-minimum antibiotic concentrations

Ampicillin, ceftriaxone, gentamicin, kanamycin, puromycin, and sulfadimidine were prepared with water as a solvent. Tetracycline, erythromycin, chloramphenicol were suspended in 70%, 100%, and 100% ethanol, respectively. Ciprofloxacin was prepared in water with HCl added drop by drop until the powder became soluble. Trimethoprim was suspended in DMSO. All antibiotics were stored in aliquots at -20ºC.

BW25113 was plated from a glycerol stock and grown for 16 hours. Three to five colonies were used to inoculate a 1 mL culture in M9 and grown for 16 hours. Samples were subsequently normalized to OD600 = 1. A 1:100 dilution was used to inoculate 50 µL cultures in M9 minimal media containing one of ten concentrations for each antibiotic (in two-fold increments) as well as controls without antibiotic. Optical density was monitored in a GENios plate reader (Tecan Group Ltd.) operating under Magellan software (version 7.2), with measurements taken every 20 minutes. The microplate reader was set to shake for 1000 seconds, with 10 seconds of shaking before measurement. The concentration selected for each antibiotic was that at which maximum OD was between 50-90% that of the control.

To determine the antibiotic resistance of the MDR bacteria to tetracycline and trimethoprim, four individual colonies were inoculated into three mL caMHB and grown overnight for 16 hours. Samples were then diluted 1:10000 in fresh caMHB, of which 45 µL was alqiuoted into a 384 well plate and supplemented with 5 µL of 10x antibiotic concentration of interest. Samples were then grown in a GENios plate reader for 24 hours using the process described above.

7.5.4 Characterizing gene knockout synergy with antibiotics

BW25113 and single gene knockout mutants were plated from glycerol stocks. Colonies from each were used to inoculate 1 mL cultures in M9 minimal media with 0.4% glucose and grown for 16 hours. Samples were then normalized to OD600=1, and were then diluted 1:100 into 50 µL cultures in M9 media containing either no antibiotic or the specified concentration of each antibiotic. Four biological replicates were included for

each condition. Optical density was monitored in a Tecan GENios as described above. The maximum OD achieved for each well was recorded and used for subsequent characterization of the nature of the interaction.

7.5.5 Characterizing CRISPR gene knockdown synergy with antibiotics

BW25113 harboring a nonsense RFP targeting sgRNA was used as a wild type control for comparing the impact of perturbations. This strain, as well as single and multiplexed gene perturbation mutants were plated from glycerol stocks, and 22 individual colonies were used to inoculate 100 µL cultures in M9 minimal media with 0.4% glucose and amp and cm selection. Cultures were grown for 16 hours, and diluted 1:100 into 100 µL cultures containing 50 ng/mL aTc and either no antibiotic or the specified concentration of each antibiotic. Optical densities were monitored in a Tecan GENios as described above, and the maximum OD achieved for each well was recorded and used for subsequent characterization of the nature of the interaction. This same process was used for characterizing synergy of Salmonella strains harboring CRISPR perturbations, except M9 media was replaced with LB media to facilitate growth.

7.5.6 Fitness assay

CRISPR strains harboring either six RFP sgRNA perturbations, multiplexed tet or trim related perturbations, or a single RFP perturbation and constitutively expressed mCherry were plated from glycerol stocks. Eight biological replicates of each strain were used to inoculate 200 µL LB cultures supplemented with amp and cm, and grown for 16 hours. Samples were then diluted 1:100 into fresh media supplemented with amp, cm,

and 50 ng/mL aTc, and grown for another 24 hours. Sample competition was then begun by mixing one µL of the mCherry strain with one µL of the strain for competition in 198 µL of the noted condition. Conditions always included LB supplemented with amp and cm, three conditions with 50 ng/mL aTc, and another three conditions without aTc. Each condition was supplemented with either no additional antibiotic, 0.25 µg/mL tetracycline, or 0.125 µg/mL trimethoprim. Two µL of each culture were used immediately to determine starting ratios of red to white cells. Cultures were then grown for 24 hours, diluted again 1:100 in fresh media, and grown for another 24 hours, after which two µL of each culture were again used to determine ending ratios of red to white cells.

Ratios were determined by plating 50 µL of 1:10000 and 1:100000 on plain LB plates. Two plate images were taken with fluorescence activation at 540 nm, one with emission filtering at 590 nm and the other without, and these images were overlaid to facilitate colony counting. Colony counts were used to determine fitness values (ω) using the standard Malthusian fitness equation, using the formula $\omega = \ln(N_{E1}\times100^2/N_{E0})/\ln(N_{C1}\times100^2/N_{C0})$ where the variables are defined as follows: “N” – CFU, “E” – experimental strain, “C” – control strain, “1” – after exposure, and “0” – before exposure.

7.5.7 Quantifying gene-antibiotic synergy

The maximum values of OD reached in the presence or absence of antibiotic and gene knockouts/perturbations were then used to determine the degree of synergy. If d_Φ is the maximum OD of BW25113 wild type in media without antibiotic, then d_X is the ODmax of the wild type with antibiotic treatment, d_Y is that of the mutant without antibiotic,

and d_{XY} is that of the mutant with antibiotic. ODmax can then be normalized as $W_X = d_X/d_\Phi$, $W_Y = d_Y/d_\Phi$, and $W_{XY} = d_{XY}/d_\Phi$. The degree of interaction synergy (S) was identified using the equation $S = W_X * W_Y - W_{XY}$. Values not statistically different from S = 0 were identified as additive interactions, while positive (or negative) values were identified as synergistic (or antagonistic). Statistical significance was determined using deviations from the null hypothesis (additive,) by performing a one-sample t-test to obtain P-values (assuming a two-tailed distribution) using the standard formula $t = (S - \mu)/(\sigma_E/\sqrt{n})$ where the variables are defined as follows: t is the test statistic, S is the average sample synergy, μ is the null hypothesis (in this case, zero), σ_S is the estimated standard deviation of synergy , and n is the number of replicates (in this case, 22).

A custom MATLAB script was written to classify interactions as additive, synergistic, or antagonistic based on the equations described above, as well as generate plots of this information. MATLAB's Statistics and Machine Learning toolbox was used to perform k-means clustering.

7.5.8 CRISPR plasmid and strain construction

CRISPR expression was driven from two plasmids, one encoding for aTc-inducible dCas9 (Addgene plasmid 44249), and another encoding for constitutively expressed sgRNA targets derived from Addgene plasmid 44251. The latter plasmid was used to create new sgRNA target plasmids by replacing the RFP-targeting sequence with new gene sequences specific to the target of interest. Unique forward primers flanked with an SpeI restriction site and encoding the new target sequence and a common reverse primer flanked with ApaI was used to PCR amplify (Phusion, New England Biolabs) new DNA

inserts, which were subsequently digested with Cutsmart SpeI and ApaI (New England Biolabs) alongside 44251 backbone. Ligation of these pieces was performed using T4 DNA Ligase (Thermo Scientific), which were subsequently transformed into electrocompetent NEB 10-β. Transformants were minipreped using Zyppy Plasmid Miniprep Kit (Zymo Research Corporation) and submitted for sequencing to confirm successful insertion (GENEWIZ). These final plasmids were co-transformed alongside the dCas9 plasmid into BW25113 to create final CRISPR individual perturbation strains.

To create multiplexed gene perturbation strains, Gibson Assembly was used to sequentially stich individual sgRNAs together. A common set of six forward and six reverse primers listed in were used to amplify sgRNAs as Gibson fragments, beginning with stitching sgRNAs together in pairs. Once the paired sgRNA plasmids were confirmed, two of the three pairs for each set were stitched together using another round of Gibson Assembly. Finally, once these four sgRNAs were confirmed, the final pair of sgRNAs were stitched together with this four-target sgRNA construct using a final round of Gibson Assembly. Final plasmid sizes were confirmed, and then transformed into BW25113 or SL1344 for experiments. All Gibson reactions were performed at 50 °C for 1 hour with T5 exonuclease (New England Biolabs), Phusion polymerase and Taq ligase (New England Biolabs). This process was used to create the 6-targeting sgRNA constructs.

7.5.9 Growth assay

To demonstrate growth curves, five biological replicates of each strain were inoculated from individual colonies into 200 µL of LB containing amp and cm in a conical

96 well microplate and grown for sixteen hours. After initial growth, two µL of each culture was used to inoculate 198 µL of fresh media containing amp, cm, and conditionally aTc or antibiotics as noted. These cultures were grown in a flat bottom 96 well microplate in a GENios reader for 24 hours as described above.

7.5.10 Statistical Analyses

The significance of synergy values was determined using a one sample t-test with a null hypothesis of $\mu_o = 0$ (i.e. no synergy or antagonism) and assuming a two-tailed distribution. Pearson correlation coefficients and their corresponding P values were calculated using linear fits with no weighting (OriginPro 9.3.226 software). All other *P* values reported were calculated using a standard two-tailed type II student's t-test. Standard deviations were estimated using appropriate propagation of error formulas excluding covariance terms.

7.5.11 HeLa culture and infection

HeLa human epithelial cells were recovered from freezer stocks in Dulbecco's Modified Eagle Medium (DMEM, Fisher Scientific), 10% Fetal Bovine Serum (FBS, Advanced, Atlanta Biologics), 50 units/mL Penicillin-Streptomycin (P/S; Fisher Scientific), and 10% dimethyl sulfoxide (DMSO, Sigma). A single freezer stock was seeded into a 96 well tissue culture treated plate (Fisher Scientific) in 100 µL, which was passaged with ~72 hours between each passage. Cells were cultured at 37 °C, 5% CO2, and controlled humidity. Cells were passaged at 80% confluency with 0.25% trypsin (HyClone). After 3-

4 passages, HeLa cells were seeded at 10,000 cells/mL and grown for another 24 hours to begin infection experiments.

The night before infection, SL1344 was inoculated from plates into 3 mL LB with amp and cm selection and grown for 16 hours. Cultures were then diluted 1:10 in LB with amp and cm and grown for another four hours, after which samples were washed three times with PBS and normalized to the same OD. HeLa cells were washed three times with Dulbecco's PBS (DPBS, Fisher Scientific) and infected in DPBS with SL1344 for 45 minutes at 37 °C, 5% CO2, and controlled humidity. HeLa cells were again washed three times with DPBS and incubated for another 75 minutes in DMEM + FBS + 100 µg/mL gentamycin to remove extracellular bacteria. HeLa cells were then washed and resuspended in fresh media (supplemented with amp, cm, 50 ng/µL aTc, and 40 µg/mL gentamycin) in either the presence or absence of 0.5 µg/mL tetracycline or 0.5 µg/mL trimethoprim and incubated for 18 hours. To perform colony forming unit analysis, HeLa cells were washed with DPBS and lysed with 30 µL of 0.1% Triton X-100 for 15 minutes at room temperature. 270 µL of PBS was added to each well and plated on plain LB agar to determine cell counts.

7.5.12 PNA design

PNA sequences were designed to bind in a centered region of the mRNA AUG start codons for the genes of interest. These 12-mer sequences consist of a KFFKFFKFFK cell-penetrating-peptide (CCP) sequence on the N' terminus, followed by an "O-linker" sequence connecting the CPP to the 12-mer nucleoside sequences with a peptide backbone. These PNA sequences were optimized to exhibit minimal off-target

effects in the BW25113 genome using a custom program described below. As *recA* exhibited an off-target around the start codon of the *ndk* gene, this PNA was redesigned to bind within the first 12 nts of the *recA* gene, beginning with the start codon. PNA sequences were also analyzed for their ability to bind to the genome of each of the clinically isolated MDR bacteria. All PNAs were found to have at least one possible target sequence in each of the MDR bacteria.

A custom Python script was used to extract the reverse complements of 12-mer nucleotide sequences centered on the mRNA AUG start codons (STC) for genes of interest. Homology was determined using the Bowtie short-read alignment tool[320] (alignment settings: -v 0 -l 12 -a), allowing for one nt mismatches or gaps within the sequence alignment. The Bedtools[321] "intersect" function was used to identify alignments that overlapped with genome features, and a custom Python script was used to parse this data and calculate the alignments' proximities to gene STCs. Off-target or homology inhibition was defined as a sequence alignment overlapping the STC of a gene that the PNA was not specifically designed to inhibit. Thermodynamic considerations for PNA sequences were screened for using a custom Python script designed to search for potential solubility and self-interference issues. The former was addressed by looking for purine stretches greater than 5 bases, a purine content of greater than 50%, or a guanine-peptide content of greater than 35%. The latter was addressed by looking for self-complementary sequences of greater than five bases.

7.5.13 PNA synergy experiments

Three biological replicates of BW25113 and MDR strains were inoculated from individual colonies in three mL caMHB and grown overnight for 16 hours. Strains were then diluted 1:10000 in fresh caMHB and aliquoted in 45 µL to a 384 well plate, to which five µL of 100 µM PNA treatment of interest (final concentration of 10 µM). Samples were grown for 24 hours in a GENios microplate reader as described above to track growth over time. Samples were normalized to blank wells and starting OD values. Values of OD were subsequently normalized to the OD of the no treatment at the same time point at which OD had reached 0.5 in the no treatment condition. Synergy values were calculated as previously described, and a one sample one tailed t test was used on the overall 48 synergy values to calculate the significance of the populations distribution in relation to the null-hypothesis of zero synergy.

7.6 Author Contributions

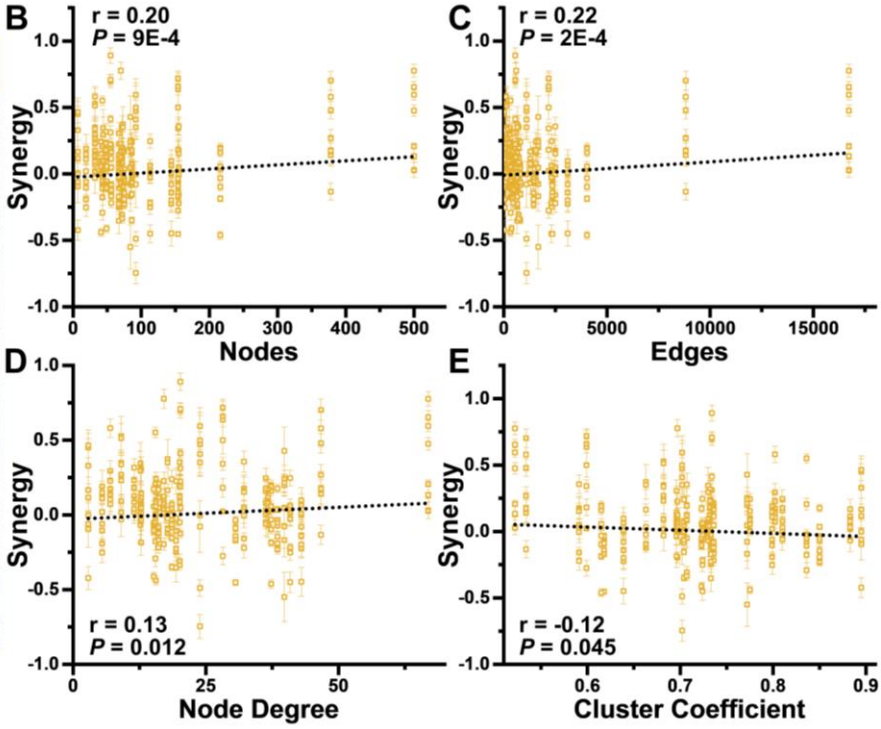
P.B.O., K.E.E., and A.C. conceived of the study. K.E.E. designed and performed all Keio knockout experiments, and quantified antibiotic concentrations for sub-MIC growth of *E. coli*. P.B.O. designed and constructed all CRISPRi strains, all PNA, and performed all CRISPRi and PNA growth experiments. K.A.E. and P.B.O. performed infection experiments. P.B.O. and K.E.E. wrote the paper.

We thank N. Madinger for providing the clinically isolated MDR bacteria used in this study. Thanks to J. Campos for her assistance in establishing infection model experiments. Thanks to T. Aunins for his assistance in screening the MDR libraries for PNA homology.

7.7 Supplementary Info

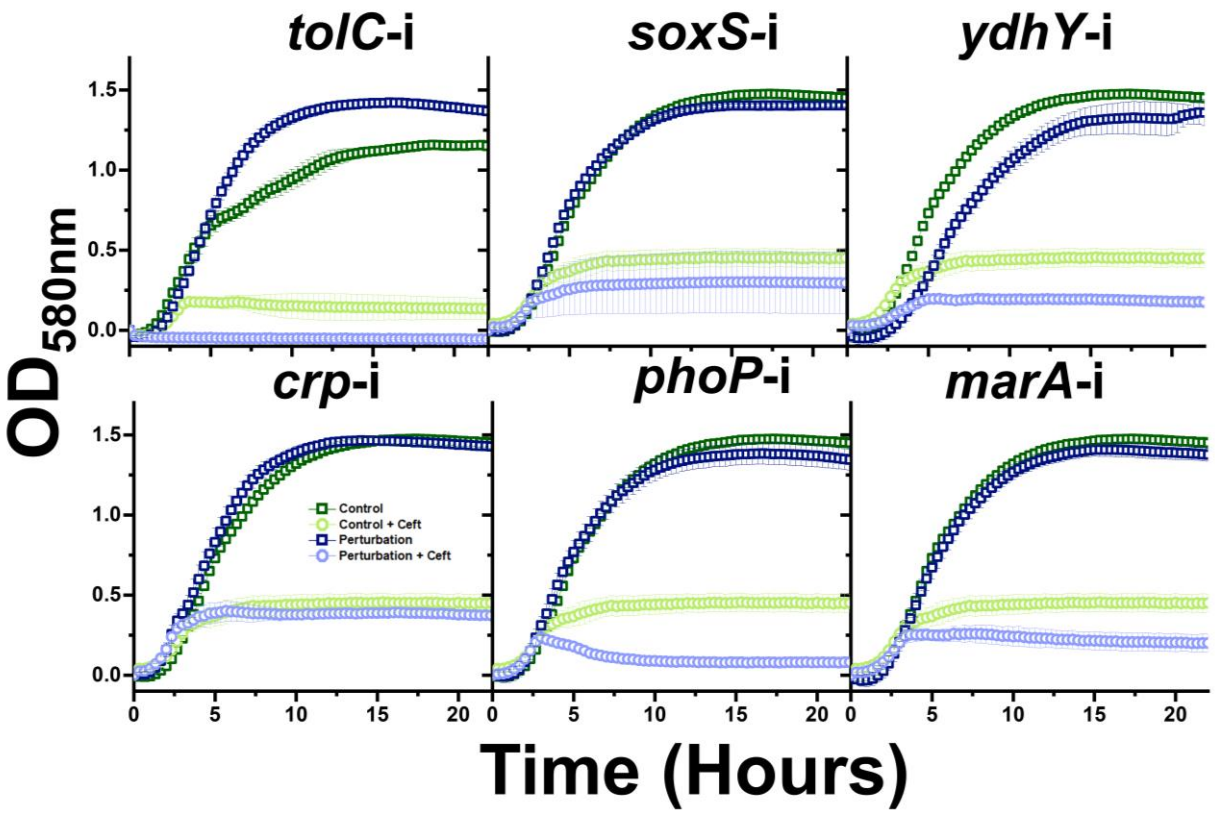
A

Gene	Nodes	Edges	Average Node Degree	Local Cluster Coefficient
ΔacrA	55	555	20.2	0.734
ΔbglG	41	319	15.6	0.775
Δcrp	216	4032	37.3	0.615
ΔcsgD	73	651	17.8	0.663
ΔcyoA	72	723	20.1	0.707
Δdam	44	279	12.7	0.730
ΔdinB	84	1670	39.8	0.772
Δfiu	43	334	15.5	0.836
ΔflhC	70	1286	36.7	0.883
ΔflhD	86	1559	36.3	0.810
Δfnr	32	112	7.0	0.802
Δfur	48	276	11.5	0.732
ΔgadX	49	372	15.2	0.723
Δhfq	154	2353	30.6	0.618
ΔmarA	56	451	16.1	0.736
ΔmutS	144	3096	43.0	0.639
ΔphoP	86	800	18.6	0.704
ΔpolB	113	2313	40.9	0.724
ΔrecA*	500	16726	66.9	0.522
Δrob	67	428	12.8	0.696
ΔrpoS	378	8817	46.7	0.534
ΔsoxS	67	650	19.4	0.733
Δtar	75	1444	38.5	0.850
ΔtolC	154	2172	28.2	0.599
Δwzc	92	1101	23.9	0.702
ΔybjG	19	52	5.5	0.799
ΔydhY	32	145	9.1	0.682
ΔydiV	155	2493	32.2	0.591
ΔyehS	7	10	2.9	0.895
ΔyjjZ	70	597	17.1	0.697



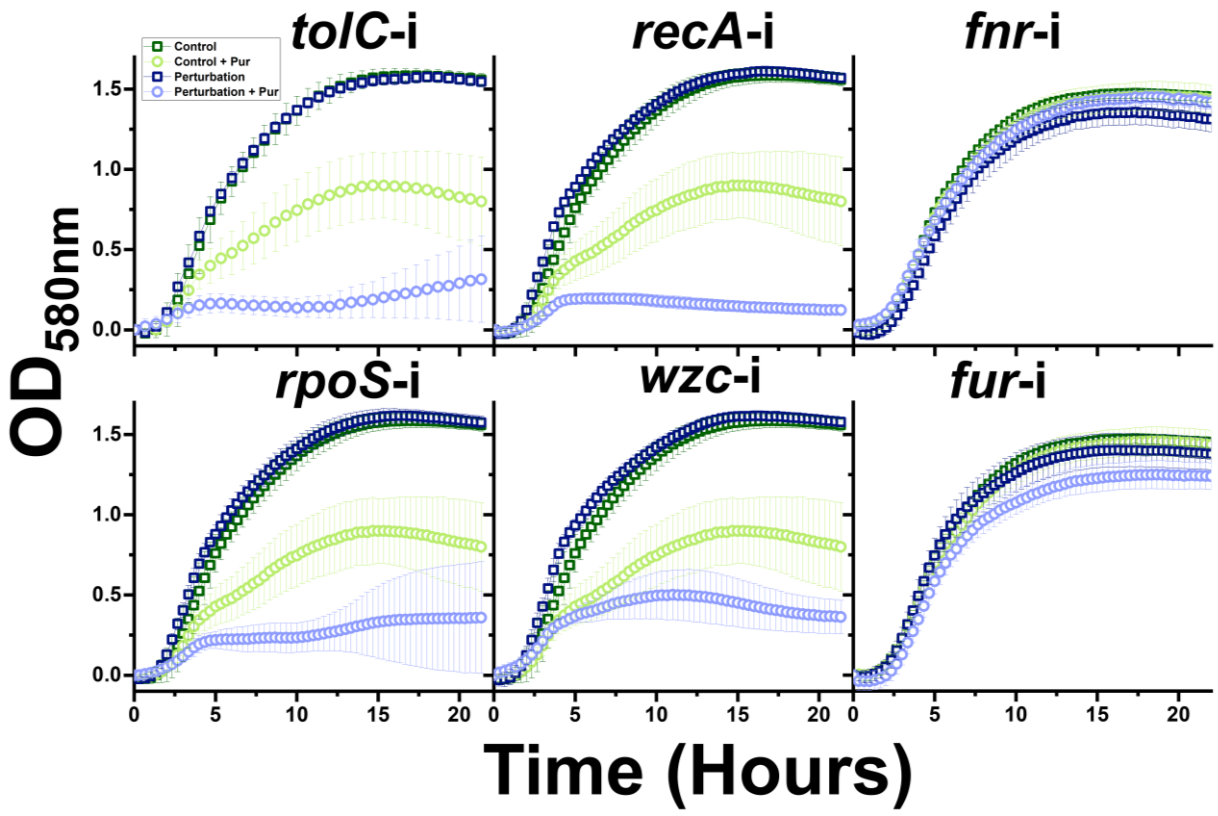
Supplementary Figure 7.1 Gene-drug synergy correlates with many characteristics of targeted gene's protein-protein interaction network

The STRING database was used to collect information on all the known protein-protein interactions that each of the thirty gene knockouts are involved in in *E. coli* MG1655. A minimum confidence score of 0.40 was used to collect information from textmining, experiments, databases, co-expression, co-occurrence, gene fusions, and gene neighborhoods. Due to output limitations, the max number of protein-protein interactions was limited to 500, affecting only *recA*. (**A**) Tabulated information on nodes (total amount of proteins in network), edges (total amount of protein-protein interactions in network), average node degree (average amount of interactions each protein is involved in), and local cluster coefficient (the "tightness" of the network indicating the degree of interaction between the network overall. This information was plotted against the degree of synergy each gene knockout exhibited with each antibiotic, leading to 270 datapoints. Significant positive correlations were identified between synergy and (**B**) gene nodes, (**C**) gene edges, (**D**) gene average node degree. This suggests that the more proteins involved in a targeted gene's interaction network, and the greater amount of interactions overall in the network, the more likely synergy is to be induced with antibiotic treatment. Conversely, a negative correlation was observed between synergy and (**E**) gene cluster coefficient, indicating that the looser the network overall, the more synergy will be induced. Y-axis error bars indicate standard deviation derived from four biological replicates



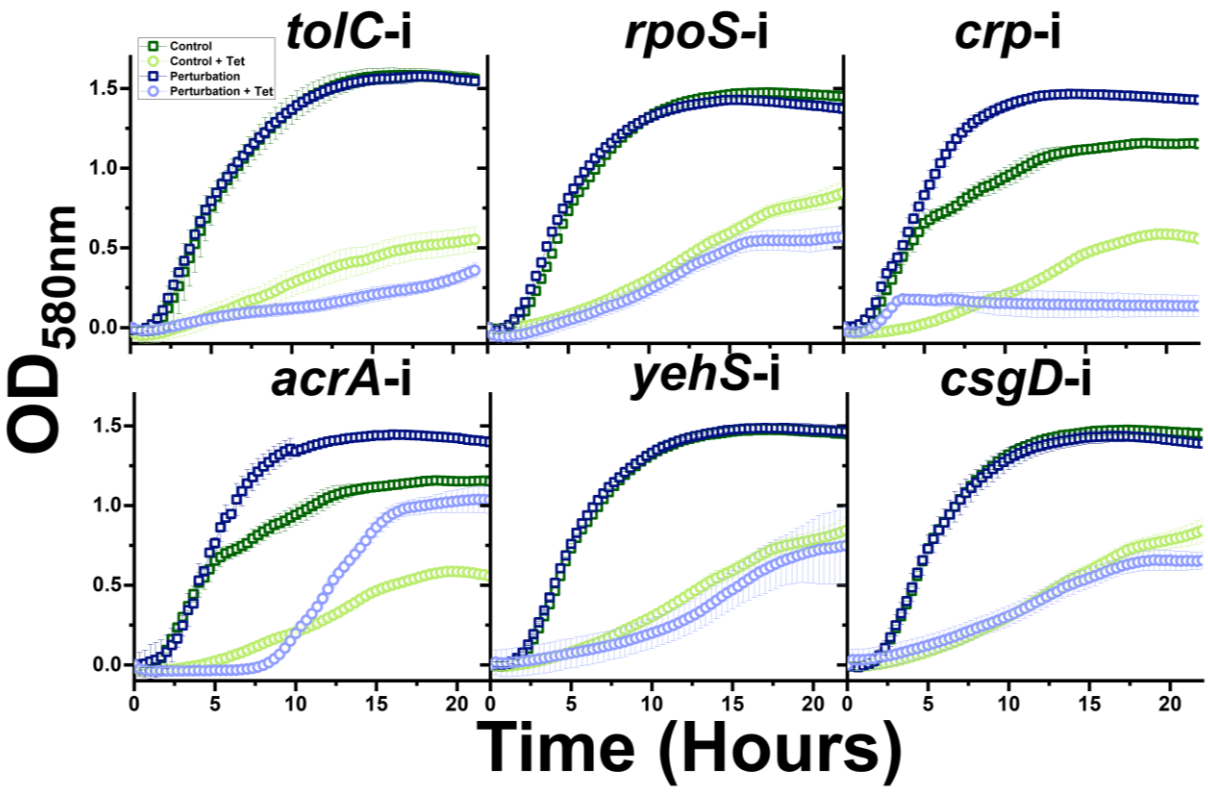
Supplementary Figure 7.2 Growth of CRISPRi strains during exposure to 2.0 µg/mL ceftriaxone in LB medium

Growth is normalized to starting ODs. Error bars represent standard deviation of four biological replicates.



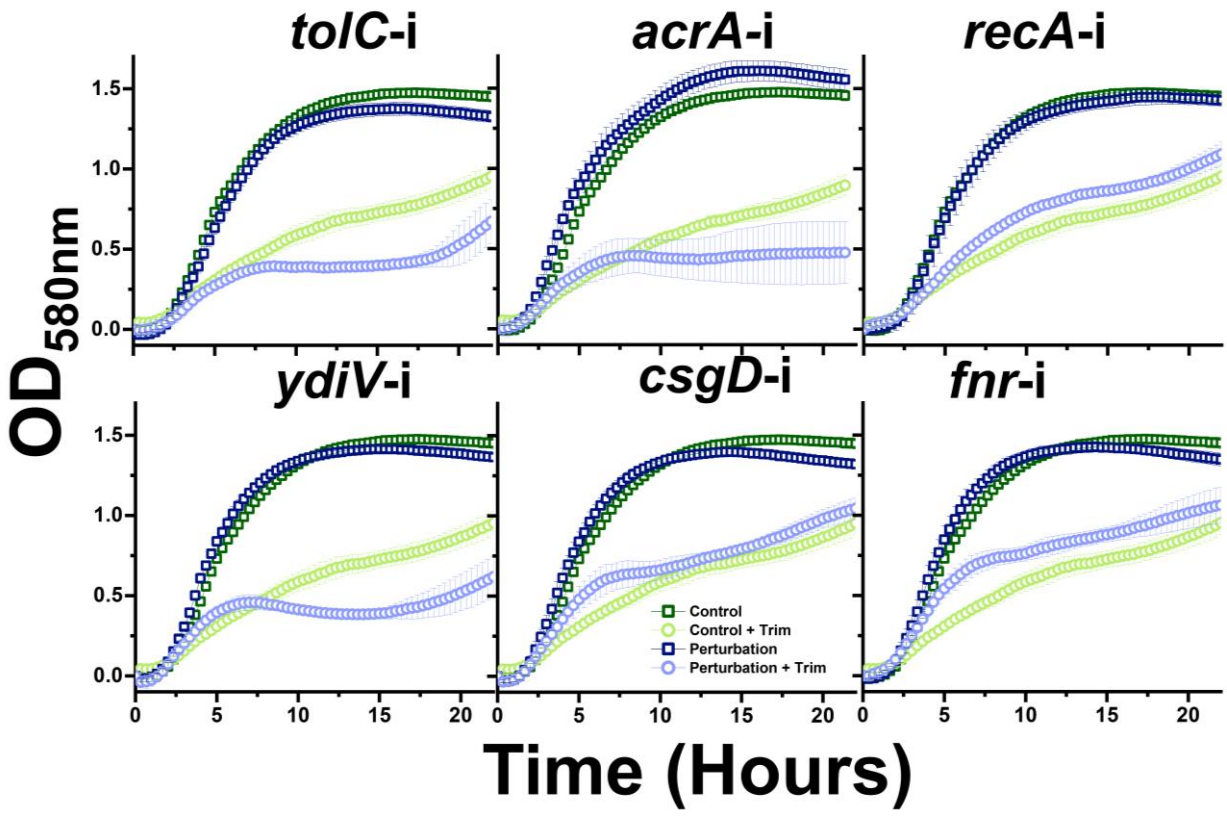
Supplementary Figure 7.3 Growth of CRISPRi strains during exposure to 50.0 µg/mL puromycin in LB medium

Growth is normalized to starting ODs. Error bars represent standard deviation of four biological replicates.



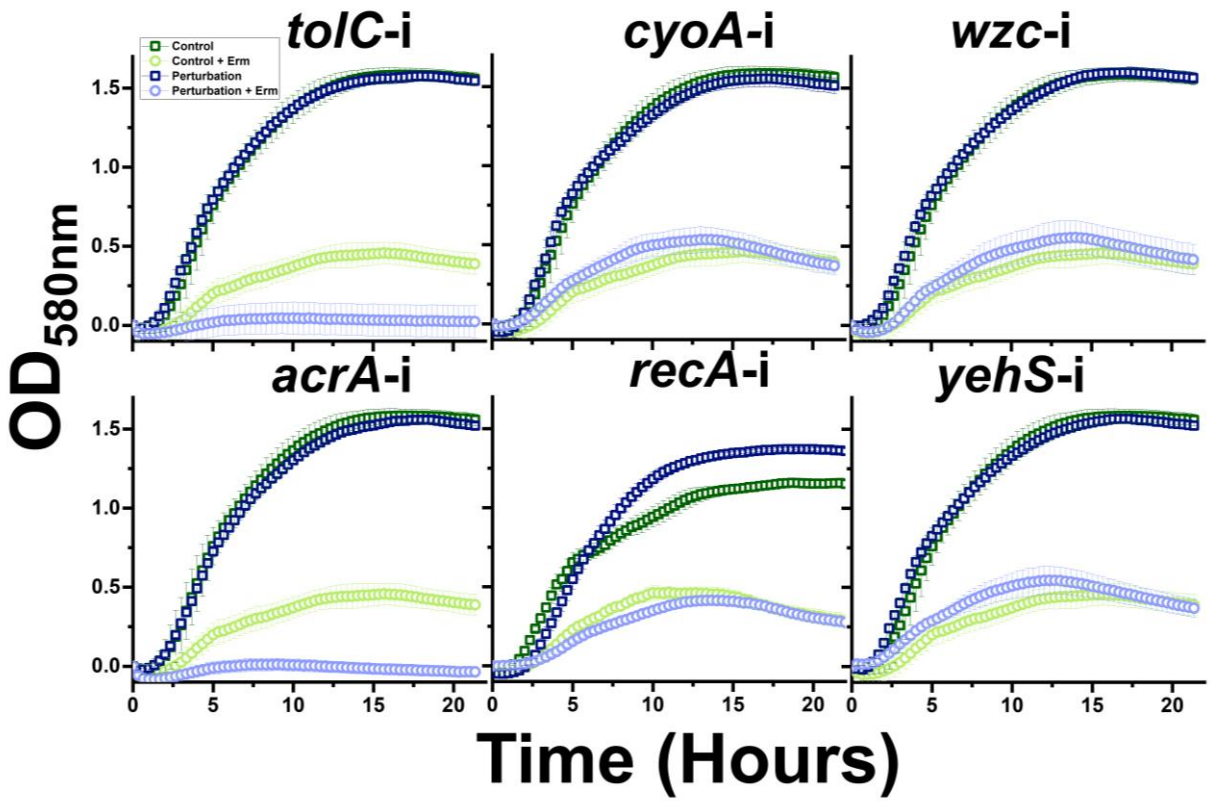
Supplementary Figure 7.4 Growth of CRISPRi strains during exposure to 0.25 µg/mL tetracycline in LB medium

Growth is normalized to starting ODs. Error bars represent standard deviation of four biological replicates.



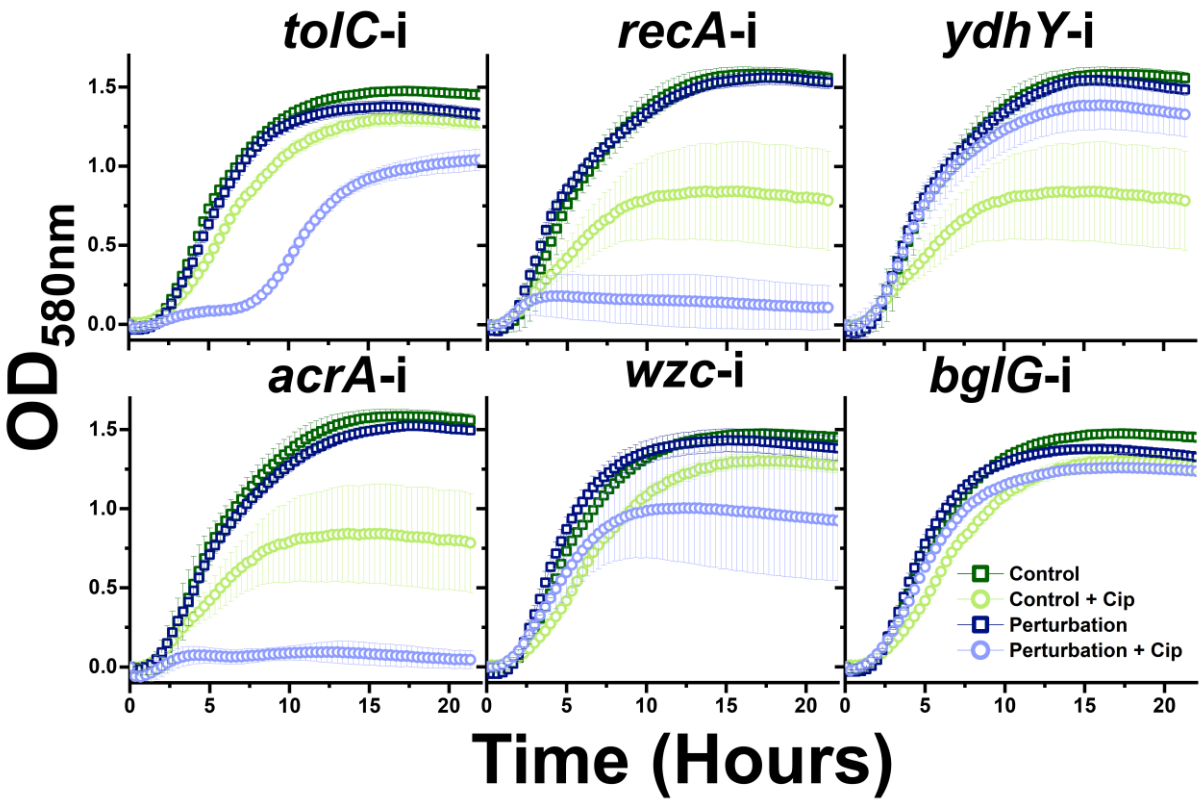
Supplementary Figure 7.5 Growth of CRISPRi strains during exposure to 0.125 µg/mL trimethoprim in LB medium

Growth is normalized to starting ODs. Error bars represent standard deviation of four biological replicates.



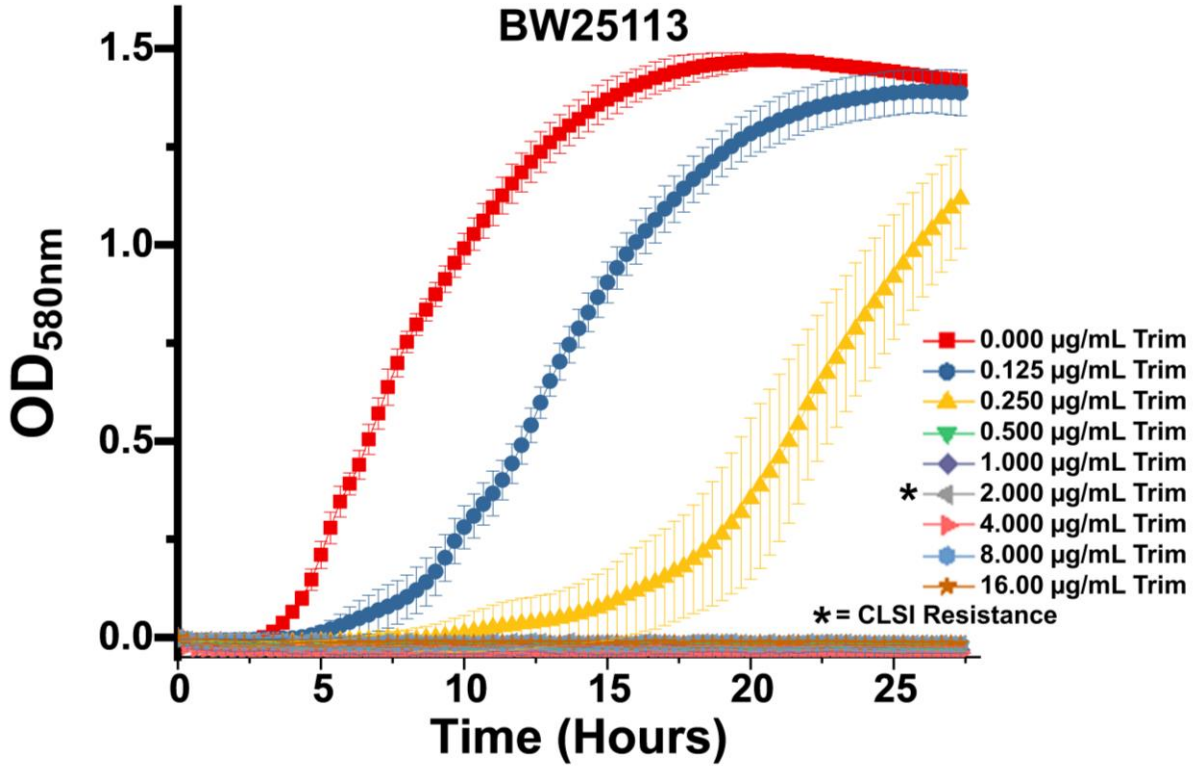
Supplementary Figure 7.6 Growth of CRISPRi strains during exposure to 50.0 µg/mL erythromycin in LB medium

Growth is normalized to starting ODs. Error bars represent standard deviation of four biological replicates.



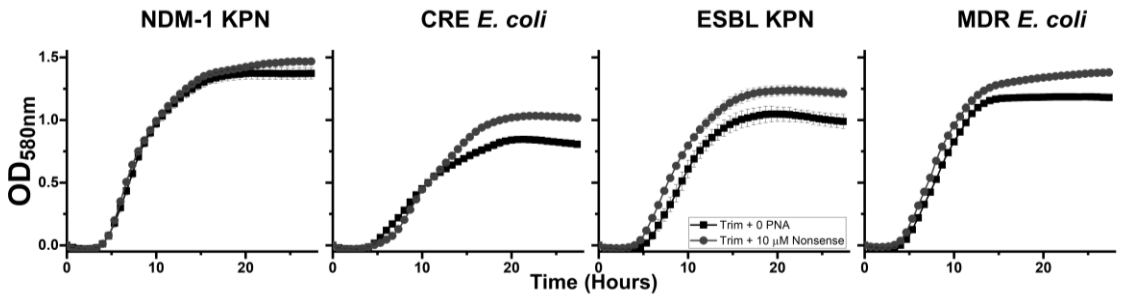
Supplementary Figure 7.7 Growth of CRISPRi strains during exposure to 0.008 µg/mL ciprofloxacin in LB medium

Growth is normalized to starting ODs. Error bars represent standard deviation of four biological replicates.



Supplementary Figure 7.8 BW25113 growth tests of trimethoprim resistance

Cultures of BW25113 were grown in caMHB for 24 hours to quantify basal *E. coli* resistance to trimethoprim. Cells were unable to survive 0.5 µg/mL trimethoprim and above, 4-fold below the CLSI breakpoint for trimethoprim resistance. Error bars represent standard deviation of biological triplicates.



Supplementary Figure 7.9 Effect of nonsense PNA on MDR growth

Growth of clinically isolated MDR bacteria during exposure to 2.0 μg/mL trimethoprim in the presence or absence of 10 μM nonsense targeting PNA. Error bars represent standard deviation of biological triplicates.

Chapter 8: Next-generation “smart” antibiotics using a holin-Cas13a kill switch

Otoupal, P. B., Cordell, W. T., Sitton, M. J., & Chatterjee, A. *In Preparation.*

8.1 Abstract

With the increasing occurrence of pathogenic multidrug-resistant bacteria and the slow discovery of new antibiotics, the need for novel antibacterial agents has increased. One new type of antibiotic proposed is a plasmid-based system that utilizes Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) enzymes delivered via bacteriophage to re-sensitize bacteria to certain antibiotics. However, often a small portion of the bacteria will mutate the introduced plasmid and survive the subsequent antibiotic treatment. In this paper, we propose to utilize a toxic “kill switch” protein (holin) and an RNA targeting CRISPR associated enzyme (Cas13a) to solve this issue. Cas13a will be used to both downregulate the kill switch and an antibiotic resistance gene and in combination, may reduce the mutational escape of pathogenic bacteria. Our results have shown the introduction of an LVA tag allows Cas13a to effectively regulate both the production of a fluorescent protein and minimize its detrimental off-target effect. This system can be adjusted to regulate a holin “kill switch” and prevent its lytic function in the future. Unfortunately, holin expression alone was not found to be effective at killing cells, suggesting the need to incorporate additional toxic elements of the native holin lysis system into our kill-switch strategy. Finally, we show that certain areas in the *E. coli*

genome are less likely to contain Protospacer Adjacent Motifs, establishing novel design criteria for CRISPR targets.

8.2 Introduction

The world is in critical need of novel antibiotics to counter the increasing number of multidrug-resistant bacterial infections, and their burgeoning threat to people across the globe. The Center for Disease Control reports that more than two million Americans are infected with multidrug-resistant strains per year, leading to at least 23,000 deaths annually in the United States alone[14]. Unfortunately, the market offers little incentive for companies to create new antibiotics, and many doctors are hesitant to prescribe antibiotics for fear of amplifying the spread of antibiotic resistance[13]. It therefore falls upon the scientific community to design and discover novel antimicrobial therapies with the issues of antibiotic resistance in mind.

A promising antibiotic approach is to turn bacteria against themselves. Specifically, it is now feasible to hijack native bacterial immune systems and cause bacteria to self-destruct. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems of bacteria, which normally target exogenous genetic elements such as viruses for degradation, have been demonstrated to serve as effective antimicrobial therapies[295]. One procedure described by Bikard et. al. utilizes bacteriophages to inject plasmids encoding for CRISPR-associated enzyme 9 (Cas9) nuclease and the single guide RNA (sgRNA) necessary to target antimicrobial resistance genes, thus re-sensitizing bacteria to a particular antibiotic[322]. Such a therapy acts in a highly sequence-specific fashion as the sgRNAs directing the nuclease are specific to a unique 20 nucleotide sequence.

Furthermore, the bacteriophages utilized are strain specific with minimal chance for off-targets effects to occur. Other approaches have focused on treatments that perturb gene expression using deactivated CRISPR enzymes targeting unique biological pathways, lessening the ability for bacteria to evolve resistance to antibiotics[19]. Bacteria have even been programmed for self-destruction, as demonstrated by Gomaa et al. who showed that the native type I-E CRISPR-Cas system of *Escherichia coli* could be exploited to eliminate itself[296].

A major limitation of these new approaches to antibiotics is the possibility of mutations arising that deactivate the bacteriophage-delivered CRISPR system[322]. If a small population of bacteria can survive subsequent antibiotic exposure by mutating the introduced plasmid to inactivate expression of the CRISPR system, then the bacteria are not reduced to an adequate level to solve the infection[322]. This potential raises a serious concern for utilizing CRISPR systems as antibiotics: can these treatments be engineered to avoid spontaneous mutation and therefore retain efficacy over time?

Here we present a novel approach to introduce selective pressure for maintenance of the CRISPR plasmid system based on co-expression and regulation of a “kill switch” gene. The kill switch is based on the native λ phage lysis system and the holin protein that enables the virus to kill bacteria. Holin functions by aggregating in the inner membrane of bacteria until a critical concentration is reached, after which the proteins associate with one another to form large holes in the inner membrane, allowing the enzyme endolysin to leak into the periplasm and destroy the bacterium[323]. In our ideal system, we incorporate constitutive expression of holin into our therapeutic plasmids. Expression of holin is concurrently down-regulated by a CRISPR system, preventing lysis.

In traditional CRISPR systems, this would require the use of deactivated Cas enzymes in order to prevent direct targeting of the plasmid DNA. However, the recently discovered Cas13a (formerly c2C2) is perfect for this application, as its nuclease activity targets only RNA while leaving the corresponding DNA intact[70]. This enzyme coupled with a guide CRISPR RNA (crRNA) can therefore be utilized in our approach to constitutively target and degrade holin RNA, preventing the T4 lysis system from killing the cell. Meanwhile, Cas13a can also degrade mRNA of an antibiotic resistance gene, making the host susceptible to a previously ineffective antibiotic treatment. Most CRISPR based antibiotics have used Cas9 or similar enzymes to regulate gene expression on a DNA level by cutting the gene. However, Cas13a is an ideal enzyme for regulating the kill switch and antibiotic in combination since it degrades mRNA while leaving the corresponding DNA intact, allowing holin to stay in place. The benefit of this approach is that if any mutation of the Cas13a system that prevents its function arises, the secondary T4 lysis system will be expressed and the bacteria will die (Fig. 1). Such an approach could be utilized in a self-replicating bacteriophage therapy as a novel antibiotic treatment.

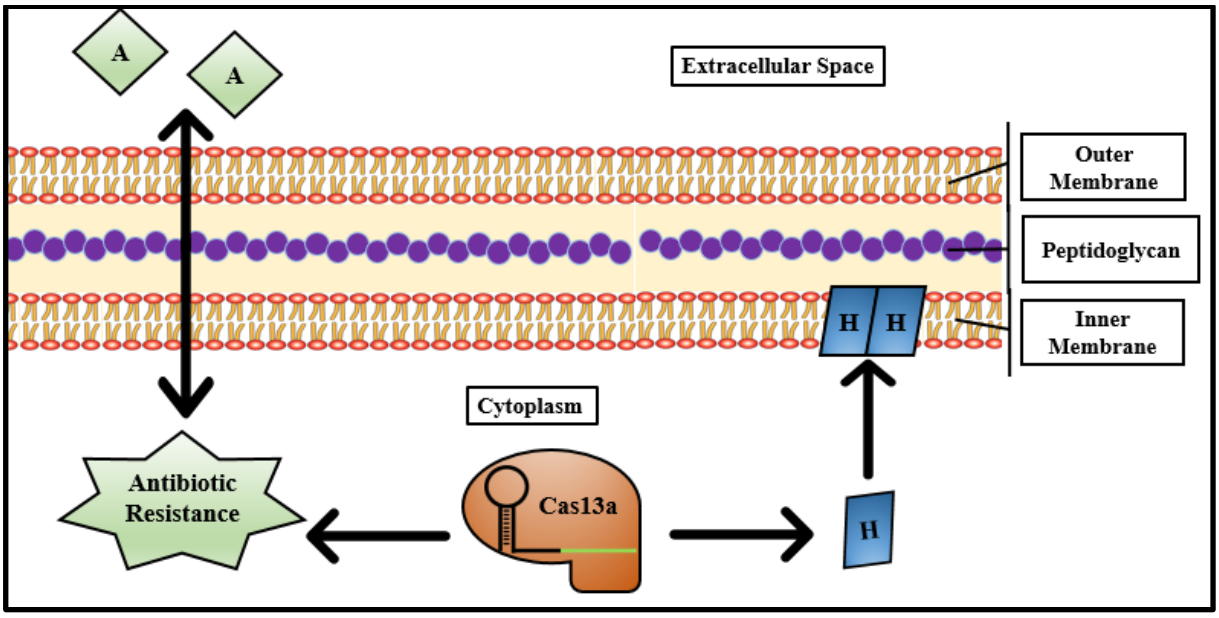


Figure 8.1 Scheme of a smart antibiotic based on Cas13a and Holin

In this designed system, selective killing of antibiotic resistant bacteria is achieved. Cas13a and Holin are constitutively expressed from a therapeutic plasmid delivered to an antibiotic resistant bacteria. Additionally, a crRNA targeting Holin mRNA and another crRNA targeting the mRNA of an antibiotic resistance gene. This prevents expression of holin protein, stopping the protein from accumulating and forming holes in the inner membrane of the target bacteria. At the same time, Cas13a reduces the expression of an antibiotic resistance gene, thereby resensitizing the bacteria to antibiotic treatment. In the case that a mutation arises in the antibiotic resistance gene target such that the Cas13a system loses its efficacy, then holin protein will continue to accumulate. This will also lead to cell death via destabilization of the cell membrane. The antimicrobial strategy is thus robust and able to maintain efficacy despite accumulation of mutations.

Targeting the holin protein is ideal for controlling cell lysis as it behaves in a step-wise fashion in controlling cell death, having no impact until a critical concentration is reached. Holin has been shown to have no negative effect on the structural integrity of the cell membrane until a hole forms[324]. Therefore, other than the metabolic burden, holin does not pose an immediate threat to the cell and thus its source plasmid could be less likely to mutate before inducing cell lysis. This dual treatment of holin and Cas13a can potentially mitigate mutations, and retain its ability to lyse cells at a higher rate than its predecessors. Furthermore, multiple crRNAs could be combined in the therapy to allow

for multiple points of therapeutic attack, circumventing issues of the target genome mutating to prevent efficacy of the treatment.

Here we lay the outline of this smart antibiotic. We show that an LVA tag mitigates the off-target effect seen in Cas13a when targeting mCherry. We see no impact on toxicity with just holin, and pose ways to fix this issue with the inclusion of more of the Lambda Phage lysis proteins.

8.3 Results

8.3.1 Holin Toxicity

The mechanism of λ bacteriophage induced lysis has received considerable attention over the past 20 years due to the λ phages' ability to replicate inside and then lyse its host bacterium. It is now accepted that the expression of the λ lysis cassette, which encodes the S, R, Rz, and Rz1 genes, is directly involved in lysing bacteria. The λ S gene encodes for holin, a protein that builds in the cytosol of the bacteria and forms large holes in the phospholipid bilayer to lyse the cell once holin has reached a critical concentration. The lambda R gene encodes for the protease endolysin which degrades peptidoglycan, causing rupture of the cell and finally Rz/Rz1 forms a complex fusing the inner and outer membrane[325]. This process is generalized in Figure 2D. Holin has the capacity to cause cell lysis independently of the rest of the lambda cassette, although efficacy is generally higher in the presence of the rest of the cassette[326,327].

We began the design of our therapy by utilizing only holin as our kill-switch operator. Holin mRNA is produced from a plasmid, which is converted into the final holin protein. This is the stage at which Cas13a will be utilized to degrade Holin, preventing the

protein from accumulating. However, in the absence of Cas13a, holin in the cytosol and dimerizes before making its way to the inner membrane. There, holin dimers collect and form oligomers or what has been termed as “death rafts”[323]. These rafts have no significant effect on the structural integrity or the proton motive force (pmf) of the cell membrane until a critical concentration of holin is reached[324]. At this step holin undergoes a change in conformation and the oligomer of holin forms a hole in the inner cell wall as shown in Fig. 2A. The pmf of the cell then drops and the cell slowly dies, leaving the outer membrane intact.

In theory, holin is the only protein that needs to be expressed in our kill-switch system. In one case of extreme overexpression of holin, it was found that bacteria lysed in as little as ten minutes[325]. The doubling time of *E. coli* is generally accepted to be 20 minutes in LB at optimum conditions[328]. Therefore, to create the most effective lysis system, holin must reach a critical concentration before the bacteria can double. For this study, this means that overexpression of holin must be achieved to minimize cell viability. The application of this system to other bacteria will depend on their doubling time in the native system. To test holin, expression from a high copy number plasmid, ColE1, under an IPTG inducible promoter was used in plasmid pHolin.

Immediate induction of the holin plasmid under 0 mM and 1 mM IPTG had no effect when comparing growth to the control (Fig. 2B). Data for inducing holin at mid-log phase also showed no effect on cell death compared to the control (Fig. 2C). This means that holin was unable to kill the bacterium with the current systems. However, when expressing the entire lambda lysis cassette including holin, endolysin, and Rz/Rz1, the effect on the cell should be complete cell lysis. The function of each of the lysis proteins

is shown in Figure 2D. Initial results expressing the entire lysis cassette show weaker, more translucent colonies even when not induced (Fig. 2E) and a clear drop in OD over time when induced (Fig. 2F). Collectively, these results indicate that although the entire λ holin-endolysin-Rz/Rz1 caused significant toxicity, holin expression alone was insufficient to induce cell death. Incorporation of the endolysin system into our kill-switch will likely be required to achieve sufficient toxicity.

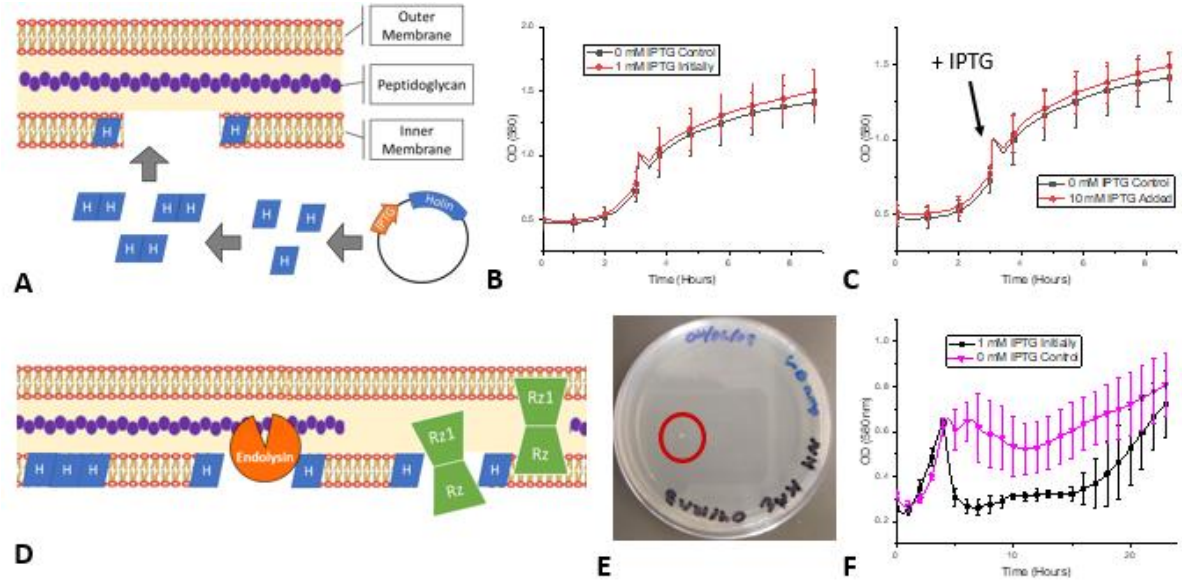


Figure 8.2 Antimicrobial activity of holin

(**A**) Expression of holin from an IPTG inducible promoter. Holin dimerizes and causes a hole in the inner membrane. (**B**) Expression of pHolin with 1mM IPTG in initially in solution compared to the control of no IPTG expression. (**C**) Expression of pHolin with induction of about 10 mM IPTG at 3 hours. (**D**) The effect of the lambda lysis cassette including holin, endolysin, and Rz/Rz1 on the phospholipid bilayer. Holin collects in the inner membrane, then forms a hole allowing endolysin to degrade the inner peptidoglycan layer, and finally allows Rz and Rz1 to form a protein complex which fuses the two membranes and lyses the bacterium. (**E**) An Agar plate showing a weak colony growing after transformation of the entire lambda lysis cassette expressed from pRG1. (**F**) DH5α-Z1 E. coli expressing the lambda lysis cassette from pRG1 over 24 hours with 1mM IPTG induction and a control of 0 mM IPTG.

8.3.2 Cas13a-LVA allows for targeted RNA degradation

With the discovery of Cas13a, a new option for regulation of gene expression has surfaced at the translation level. Cas13a, previously known as C2C2, is isolated from the bacterium *Laptotrichia shahii*. For Cas13a to function, it must associate with a single-component crRNA. Cas13a is able to process this crRNA itself, a unique feature that few other CRISPR systems share. The crRNA has secondary RNA structure and target sequence that, once associated with Cas13a, provides an RNA target. The homologous crRNA target sequence binds to the target RNA, and Cas13a cuts the target.

One known disadvantage of Cas13a protein is that after degrading its intended target, Cas13a indiscriminately degrades all other RNA around[48]. This effect can slow or even halt bacterial growth. Expression of a wild-type Cas13a would cause a generally negative effect and is likely to lead to faster mutation in pathogenic systems. Therefore, it is necessary to mitigate the off-target effect of Cas13a. One possible solution is a protein degradation tag which can be used to degrade Cas13a before it can significantly affect cell growth. Protein degradation tags already exist inside bacteria to target and degrade problematic proteins. The protein degradation tag is a short peptide sequence that can be added to proteins to decrease their half-life, or the average time it takes for half of the proteins to degrade. The one selected for this study, an ssrA tag, is temperature sensitive and has a half-life of about 13 minutes at 37℃ and 41 minutes at 25℃[329].

We tagged Cas13a with such a protein degradation tag (referred to as LVA), and investigated how this affected its ability to regulate expression of a mCherry red fluorescent protein. This tag was attached to both the N-terminus and C-terminus in the case that one would prove ineffective. The control used for these experiments was a

sgRNA which targets a red fluorescent protein. sgRNA should not associate with Cas13a since it does not have the correct secondary structure, thus it should have no effect on the system.

Initial results show that targeting mCherry with Cas13a resulted in substantially lower expression of mCherry, indicating that the Cas13a crRNA system effectively targets and degrades the target mRNA (Fig. 3A). Attaching the LVA tag to either the N-terminus or C-terminus resulted in a decrease in mCherry degradation capacity but N-terminus LVA tag appeared to be more effective in this regard. Analyzing the N-terminus, the best control appeared to be the sgRNA 006 expressed at the same level as the crRNA 001 target. Figure 4B shows the expression of Cas13a-LVA with mCherry protein and the crRNA 001 targeting mCherry, and the nonsense sgRNA at 1mM IPTG and 50ng/ml aTc. We found an increase in expression of mCherry relative to plain Cas13a. This indicates that the tag effectively prevented excessive off-target mRNA degradation. Furthermore, expression of mCherry was in between the Cas13a mCherry targeting strain, and the Cas13a-LVA nonsense targeting strain. This indicates that a statistically significant reduction in fluorescence occurred when targeting mCherry with Cas13a-LVA. When the final fluorescent points were normalized to the optical density measured at the end of the experiment, the difference in means between Cas13a-001 and Cas13a-006 were statistically significant with an alpha of .005.

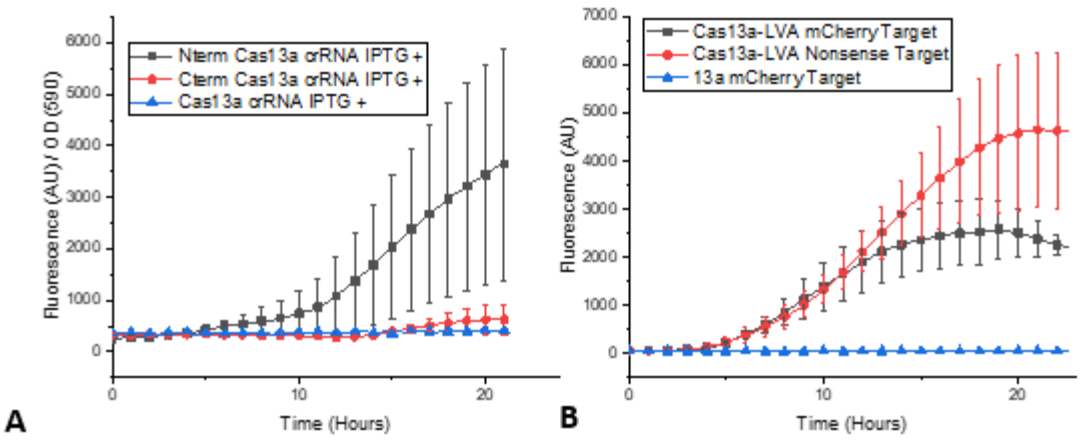


Figure 8.3 LVA tag reduces activity of Cas13a

(**A**) Expression of Cas13a with an LVA tag on the C-terminus and N-terminus of the enzyme at 1mM IPTG driving expression of mCherry and 1ng/ml aTc driving expression of Cas13. The y-axis is fluorescence normalized to by the optical density of the culture. (**B**) Expression of mCherry in strains harboring mCherry-targeting Cas13a-LVA on the N-terminus and Cas13a. Under aTc induction of 50ng/ml, expression of mCherry from a Lac promotor with a concentration of 1mM IPTG, and constitutive expression of the mCherry crRNA target and control sgRNA.

8.3.3 Design principles for crRNA targets

Cas13a can be designed to target not only antibiotic resistance genes, but also virtually any gene in the genome. This could be useful in making more broad-spectrum antibiotics based on Cas13a. However, it would be useful to know particular genomic locations that are better for targeting than others. One such technique might be to target regulons, which consist of multiple similar sequences throughout the genome that are targeted by transcription factors to modulate transcription rates. Since these sequences are common and repeated throughout the genome, we hypothesized that selective pressures throughout evolution have driven bacteria to protect these areas more than others, particularly from their own CRISPR systems. Specifically, we wondered whether or not these transcription factor regions are less likely to exhibit natural Protospacer Adjacent Motifs (PAM). Bacteria may have evolved to bias these transcription factor

regions to have less PAM binding sites to prevent their own CRISPR systems from binding in these highly conserved regions.

To investigate this, we examined the regulons of four transcription factors in *E. coli*: Fis, H-NS, Lrp, and IHF. We used the Regulon Database to extract the known transcription factor binding sites of each of these regulators, as well as the ten nucleotides upstream of these sequences. We then scanned this ten base-pair region for any PAM site that the type I-E CRISPR system of *E. coli* utilizes. This includes 5' ATG, AAG, AGG, GAG, or TAG from the non-target strand[330]. We counted the sum of all PAM sites that appeared in this region. We next constructed an equal amount of random 10nt long DNA sequences, and calculated the average number of PAM sites identified in these sequences. We did this for six random data sets, and found the average count of PAM sequences in this data set. We then examined the normal distribution of these data sets and found the statistical chance of the actual number of PAMs in the regulon of falling within this range. We found a statistically significant reduction in PAM motifs in all of these regulons, with the exception of the lrp regulon (Table 1).

Table 8.1 PAM representation in *E. coli* regulons

The total number of PAM sites in the 10 nts upstream of specific transcription factor binding sites compared to PAM distribution in six random datasets of the same size.

Regulon	Genes in Regulon	PAM Sites in Regulon	PAM Sites in Random Dataset	Significance (α %)
Fis	227	116	142 ± 14	1.7
H-NS	53	22	33 ± 5.7	2.8
Lrp	84	49	53 ± 7.0	29
IHF	101	38	63 ± 11	0.12

This data shows that transcription factor binding sites of certain regulons are statistically less likely to contain PAM motifs. This data suggests that evolutionary pressure has driven these regions to avoid containing PAM motifs in order to prevent spontaneous DNA degradation.

8.4 Discussion

The length of the pHolin plasmid was confirmed by restriction digest and the sequence showed roughly that the plasmid likely contained holin (Fig. S1). One possibility for the failure of the holin to cause lysis by itself could be insignificant holin production that was unable to keep up with cell division, meaning the threshold concentration was never reached. Another explanation could be that, due to the leaky expression of the IPTG inducible promoter, a weak expression of holin could have caused an undesirable effect on the cell but did not cause cell death, and thus was mutated over time. Further analysis of holin production and possible overexpression should be considered in future experimentation.

Addition of an LVA tag to Cas13a has decreased the negative effect on this system and also allowed Cas13a to retain its ability to target select proteins as shown by the decrease in fluorescence when targeting mCherry. It is also clear that the spread of the error bars grows for the cultures over time. This is typical of biological systems because separate bacterial populations, even when stated from with bacterial replicates, can have different phenotypes form and begin to diverge. This could be minimized with the inclusion of more bacterial replicates in future experiments.

Although creating a dual plasmid system with two adjustable promotors has allowed some amount of tuning for the two proteins produced, the leaky expression has made it difficult to find an accurate baseline of expression. There appears to be no difference between activation with 1 ng/ml and no induction for the expression of Cas13a. Furthermore, long-term cultures still have leaky expression of both Cas13a under the single tet promotor, and mCherry under the IPTG inducible promoter. This is likely why Cas13a expressed with the mCherry target crRNA cannot grow well even without any induction.

There remains a number of avenues for optimizing and completing this kill switch system. If the expression of holin insufficient for cell death at expression levels possible in this system, then the inclusion of multiple proteins from the lambda lysis cassette could improve bacterial control. The original lambda lysis cassette included holin, holin's inhibitor, endolysin, and Rz/Rz1 or the spannin proteins. The holin inhibitor, also known as anti-holin, should not be included to keep lysis time low. Endolysin, a protease which degrades the peptidoglycan protein layered between the inner and outer membrane, can cause full cell lysis when external stress is placed on the cell. The function of the spannin proteins is to fuse the inner and outer membranes after holin and endolysin have completed their function. Including all three proteins would create a system capable of lysing the cells fully without any external pressure. Tests of this whole system, including anti-holin, have shown a significant reduction in OD and thus cell viability. Regarding the smart antibiotic system, Cas13a would only have to target holin to prevent the function of endolysin and spannin, leaving a relatively simple system in place.

Since Cas13a-LVA has been shown to regulate mCherry, testing a system where holin is regulated could prove the efficacy of the system purposed in this paper. Repeating experiments to prove the viability of the Cas13a-LVA system with a finer spread of aTc activation concentrations will elucidate how much control we can have over this system. Finally, improvement of the leaky expression could be achieved with a second aTc inducible promotor and a different IPTG inducible promoter.

8.5 Materials

8.5.1 Strains and culture conditions.

Escherichia coli (*E. coli*) strain Neb 10β was used for all cloning. Strains MG1655 and DH5α-Z1 were used for all experiments. Media used was created from a powdered Lennox lysogeny broth (LB, Sigma-Aldrich). Chloramphenicol (cm) and ampicillin (amp) were used as selective markers for plasmids at 35 µg/mL and 100 µg/mL, respectively, on both liquid and solid media. The Isopropyl β-D-1-thiogalactopyranoside (IPTG) concentration used in solution was 1-10 mM from a 1 M stock. The maximum anhydrotetracycline (aTc) concentration was 50 ng/mL in solution.

8.5.2 96 well plate experiment setup.

Typical plate experiments began with a single colony of the relevant strains containing desired plasmids grown overnight in selective media and then made into a glycerol stock the following day by adding 500 µL of culture to 500 µL of 50% filter sterilized glycerol in water. These freezer stocks were used to streak a selective agar plate of colonies to be used for bacterial replicates. Typically, five colonies were taken

form the previous plate and either grown in a 96 well plate in 200 µL of LB or a 5 Ml LB culture overnight in appropriate antibiotic conditions and was used the following day for experimental cultures. At least three replicates were used and each bacterial replicate began with a 1 to 100 dilution of the overnight culture. IPTG and/or aTc was added to each well and the plate was read using a Magellan Tecan plate reader in flat-bottom plates. Cycles were measured at 20-minute intervals with 1000 seconds of shaking between measurements and 10 seconds right before measurement. Optical density (OD) measurements were taken at an absorbance of 590nm or 580nm while fluorescent measurements were taken at an excitation of 580nm and an emission spectrum of 620nm. Data was taken in 24-hour intervals unless otherwise mentioned. Non-experimental wells were filled with sterile water, or extra LB. If both OD and fluorescent measurements were needed, 198uL of the appropriate media conditions were inoculated with 2 uL of culture were mixed and then split evenly between OD and fluorescent plates for a total volume of 100 uL of culture in each experimental well.

8.5.3 Designing Cas13a targets.

crRNAs require a 5'UTR repeat sequence needed for the secondary structure needed to associate with the Cas13a. The 5'UTR is upstream of the desired target sequence of 28 nucleotides (nt)[72]. crRNA targets were designed using the methods described by Gootenberg et. al. The produced RNA from the plasmid bottom 3' to 5' strand must match the RNA target from the top strand of the target DNA 5' to 3'[73].

8.5.4 Plasmid construction.

Restriction digest of plasmid components was done using the suggested protocol form with New England Biosciences or Thermo Fisher depending on the enzymes used. Ligation of plasmid pieces was completed using the protocol from the supplier, Thermo Fisher. Gibson assembly followed NEB master mix conditions and protocol. The master mix was created in lab from its individual components. DNA segments were ordered from Twist Biosciences and cloned into plasmids using the methods mentioned above. Plasmids used in these experiments are summarized in Figure 8.4.

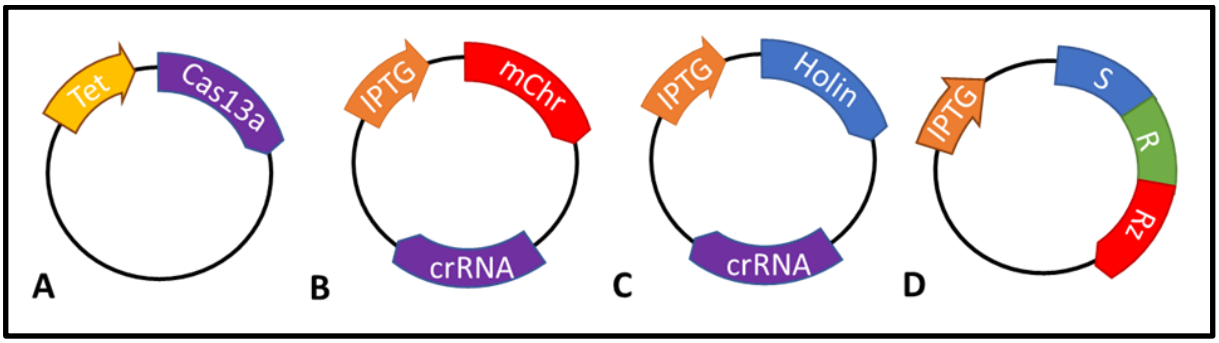


Figure 8.4 Plasmid schematics

(A) Plasmid pCas13a expresses Cas13a or Cas13a-LVA from a tetracycline (aTc) inducible promoter. (B) Plasmid pRFPcrRNA expresses mCherry from an IPTG inducible promotor and crRNA with a constitutive expression promotor, J23119. (C) Plasmid pHolin expresses mCherry from an IPTG inducible promotor and crRNA with a constitutive expression promotor, J23119. (D) Plasmid pRG1 expresses the lambda lysis cassette including lambda S gene encoding for holin and antiholin, the lambda R gene encoding for endolysin, and the lambda Rz/Rz1 gene encoding for Rz/Rz1 under an IPTG inducible promoter.

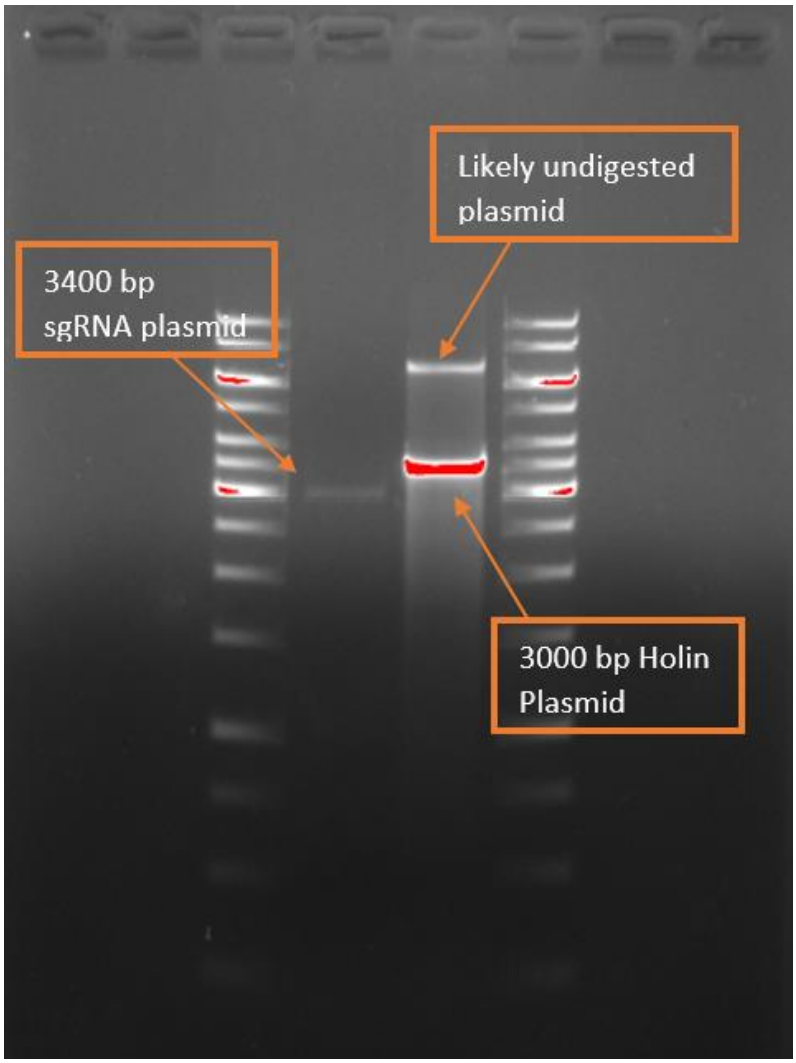
8.6 Author Contributions

W.T.C. and P.B.O. devised and planned the study, performed all experiments, and wrote the manuscript. W.T.C., M.J.S., and P.B.O. constructed all plasmids. P.B.O. and A.C. revised the manuscript.

8.7 Supplementary Info

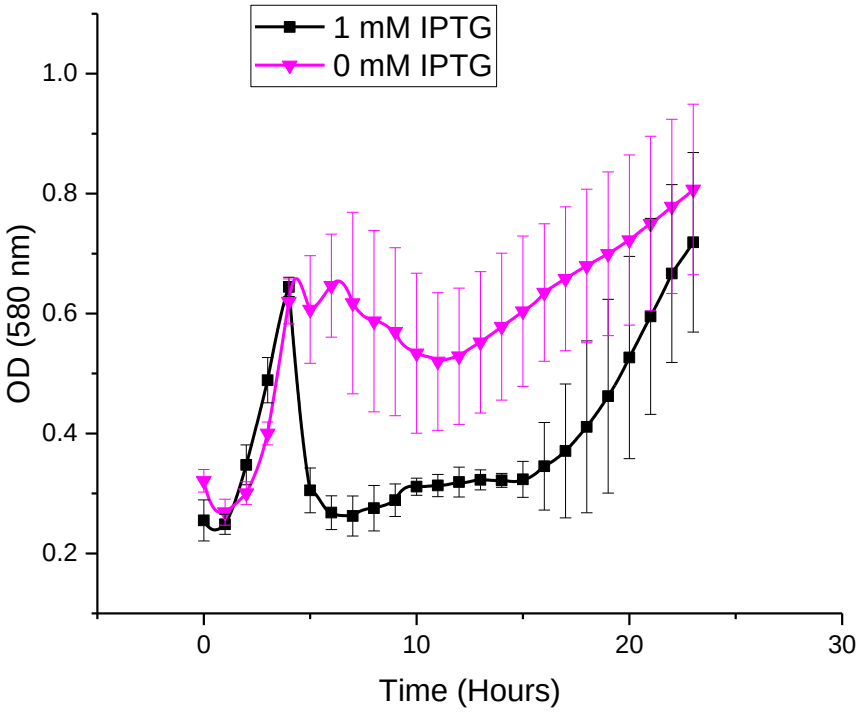
8.7.1 Supplementary Figures

.



Supplementary Figure 8.1 Holin plasmid size confirmation

Digestion comparison of a comparative sgRNA plasmid (3400 bp) in lane 2 compared to the digestion of the crRNA Holin plasmid (3000 bp) with BamHI. The two bands in the third lane likely represent undigested plasmid since no DNA of a size greater than 3000 existed, and the correct digested plasmid of size 3000bp.



Supplementary Figure 8.2 pRG1 toxicity

Expression of plasmid pRG1 provided and its impact on bacterial growth. Here a culture of DH5az1 which represses expression of the lysis proteins, is activated at a concentration of 1 mM IPTG in solution at the beginning of culturing. There is a clear drop at hour 4, signifying lysing of the culture.

Chapter 9: Engineering sequence-specific control over bacterial translation rates using deactivated Cas13a fused to IF1

Otoupal, P. B., Sitton, M. J., Cordell, W. T., & Chatterjee, A. *In Preparation.*

9.1 Abstract

Despite the enormous progress made in the synthetic biology field, there has yet been a tool developed to increase translation rates. This tool would benefit numerous areas of research, ranging from therapeutics development to metabolic engineering. Here we present the development of a CRISPR-based platform for increasing bacterial translation rates. This is based on the fusion of either deactivated Cas9 (dCas9) or dCas13a to the bacterial translation initiation factor 1. We describe our current progress in its ability to modulate translation rates in an effective, specific, and tunable manner. We will show results from several experiments targeting a fluorescent protein, mCherry, and the genome-based lacZ gene. Our experiments demonstrate significantly promising results, as in certain conditions CRISPR-IF1 fusions can lead to many fold increase in gene expression. However, further characterization is necessary in order to ensure that these increases are due to effects at the translational level. Additionally, we explore the design principles for optimizing gene targeting with these fusion proteins to increase translation rates. Interestingly, our greatest success with both fusion proteins resulted when they were targeted in the middle of open reading frames. Taken together, these results show great potential for the use of this approach as a novel platform for altering bacterial translation rates, and present guidance for how such a platform should be designed.

9.2 Introduction

As the synthetic biology continues to advance at breakneck speed, there is increasing interest in not only editing genes, but also regulating their expression. This technology is of such great interest because it provides researchers with the ability to modulate expression in order to parse out gene networks. A better understanding of how genes interact in both eukaryotes and prokaryotes can benefit numerous fields, including better understanding of healthy and diseased states in humans to provide new targets for novel therapeutics and a more efficient exploitation of prokaryotes to produce products like biofuels, pharmaceuticals, and food products. Furthermore, modulating gene expression, in lieu of gene editing, is a much safer option for therapeutics and clinical applications. So far, research into modulating gene expression has focused around activating or repressing transcription rates of key genes using DNA-binding proteins, like dCas9, which is typically linked to a molecule that either increases or decreases rates of transcription.

The discovery of clustered, regularly interspaced, short palindromic repeat (CRISPR) systems and associated (Cas) proteins in bacterial genomes has allowed researchers to manipulate and modify genomes with greater efficiency and accuracy than previously achieved. The native CRISPR system acts as an immune system for bacteria and archaea[37]. For regulating gene expression, however, researchers use deactivated forms of the Cas proteins. These deactivated forms of the proteins are no longer able to cleave, but retain their ability to target and lock on to specific sequences. In this way, researchers are able to use these deactivated forms of Cas proteins to regulate gene expression, both through activating and repressing expression. For example, previous

studies have shown that gene expression is able to be repressed by targeting dCas9 to the promotor region of a target gene, thus preventing RNA polymerase from binding and initiating transcription[44]. Similarly, modified dCas9 fusion proteins such as dCas9-ω[45] and dCas9-VP64[331] have been shown to successfully increase expression via recruitment of RNA polymerase to specific DNA elements.

However, there are several disadvantages to targeting transcription in order to modulate gene expression. For one, there are numerous proteins, like transcription factors, that are required for the initiation and execution of transcription. If one were to target the transcription step using DNA-binding proteins, dCas9, for example, would have to compete with these other necessary proteins in order to bind to critical regions, like the promoter. The effect of these activator and repressor proteins are thus weakened because of the competition with the native regulatory system. Furthermore, targeting DNA elements is much more difficult in eukaryotic systems, where the protein must first enter into the nuclease. And finally, regulating gene expression at the DNA level means that entire operons must be modified in tandem; repression of an upstream gene will inherently reduce expression of all downstream genes, making it difficult to tune individual genes, especially in complicated structural networks.

Targeting the translation step of the gene expression process, however, would likely provide a much more desirable terrain to work with. Translation is performed at the individual RNA level with very few protein factor interference, and thus translation-engineering strategies would have much less competition to bind to its target. While there are tools that can repress gene expression at the translation level, like small interfering

RNA (siRNA)[332] and peptide nucleic acids (PNAs)[94], no tools have yet been developed that can increase the rates of translation.

Here we report our attempts to develop a CRISPR-based tool for programmably increasing translation rates in bacteria. We fuse CRISPR enzymes dCas9 and dCas13a to the bacterial initiation factor 1, in order to increase the local concentration of the latter protein in particular regions of the cell near particular mRNAs. We describe both the successes and the failures of this approach, demonstrating that there is indeed significant potential for this strategy to increase translation rates but that further characterization is necessary to reliably ensure translation activation. This work paves the way for a novel translation-engineering tool that will provide a more effective tool for researchers to manipulate gene expression in prokaryotes, allowing for studies of gene expression networks in numerous fields.

9.3 Results

9.3.1 Design of dCas9-infA fusion protein

We began this work by attempting to fuse initiation factor 1 (IF1) to the canonical deactivated CRISPR enzyme, dCas9. IF1 plays two crucial roles in the translation process[333]: 1) blocking the A-site of the 30S ribosomal subunit, thereby forcing the initiator tRNA to bind to the P-site, which begins translation and 2) enhancing association and dissociation rates of the 70S ribosomal subunit via interactions with IF2 and IF3. Studies have shown that cells without IF1 have slow growth rates and short polysomes[334]. IF1 was chosen for the fusion protein because of the prominent role it plays in the initiation of translation and because it is the smallest of the three translation initiation factors, with a

molecular mass of 8.2 kDa and a gene length of only 216 nucleotides. We hypothesized that having a greater concentration of IF1 near the ribosome binding site of target RNA would encourage more ribosomes to bind there, and thus create more of the desired protein

While dCas9 is unable to target RNA normally, work from the Doudna lab has shown that the use of single-stranded DNA complementary to the target RNA containing a Protospacer Adjacent Motif (PAM) sequence can enable dCas9 targeting of RNA[335]. We hypothesized that using the single-stranded DNA sequences (called PAMmers) alongside a dCas9 fused to initiation factor 1, we could increase the local concentration of ribosomes near a target mRNA and therefore increase bacterial translation rates (Fig. 1). To test this, we built a dCas9 construct with the entire *infA* gene from *E. coli* attached to the C-terminus using a rigid EAAKEAAK linker. We also built eight various sgRNAs targeting the mRNA of mCherry (I1-I8). The binding sites of these eight targets are shown in Figure 1, and include two sequences that bind to areas with an ideal PAM site (NGG), four that bind to a slightly mismatched PAM site (NGA), and two that bind to a PAM site with no ability for dCas9 binding (NRR). The first two sites allow for DNA targeting, while the next four allow for DNA targeting with a slightly reduced efficacy. The last two have no DNA targeting efficacy, and can therefore be utilized alongside a PAMmer.

As seen in previous works to increase and decrease rates of transcription, the positioning of the fusion protein is critically important to the potency of the construct[45]. If the construct is positioned too close to critical regions where other machinery must bind to carry out transcription or translation, like the promoter region in transcription or the ribosome binding site in translation, the construct may prevent binding of this crucial

machinery and thus decrease the rate of expression. In turn, if the construct is too far away from the target DNA or RNA, the fusion protein may show little to no effect. Thus, we imagined the existence of a “sweet spot” for the dCas9-IF1 to have maximum efficacy. The similar CRISPR fusion protein dCas9-ω was found to have maximum effect when targeted ~80-100 nucleotides upstream of the transcription start site[45]. We held this in mind in choosing our target positions.

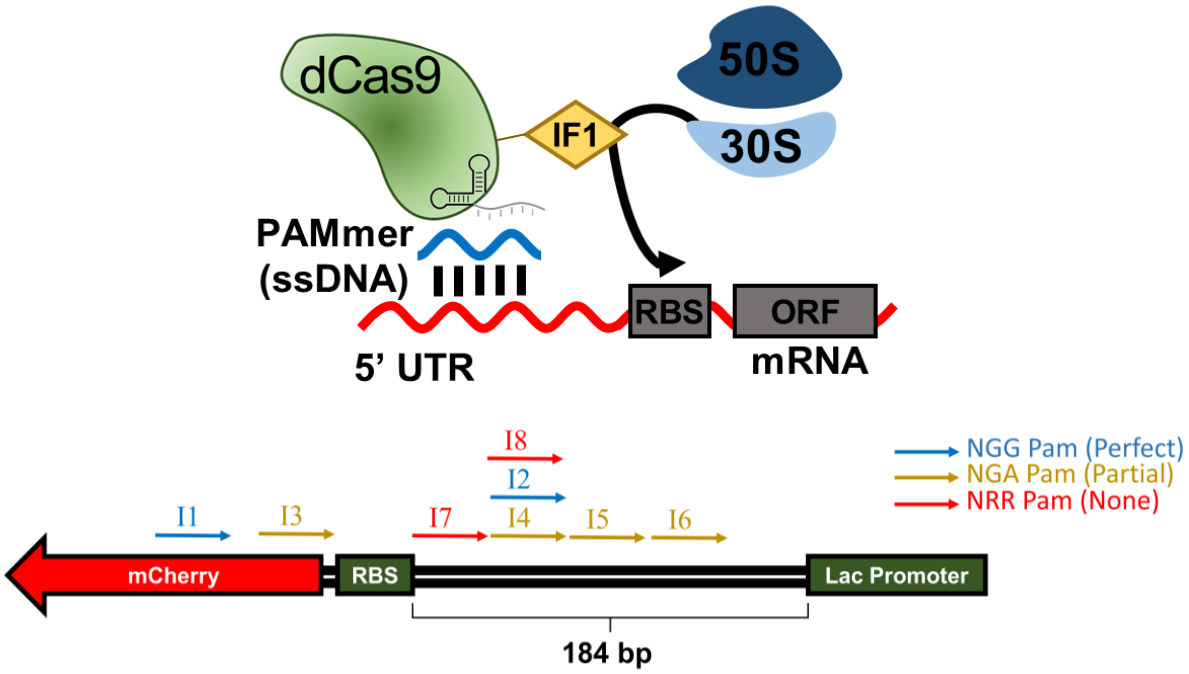


Figure 9.1 Design of dCas9-IF1 to increase mRNA translation rates

A schematic depicting how dCas9-IF1 can increase bacterial mRNA translation rates. Here, dCas9 is fused to initiation factor 1 by attaching the *infA* gene from *E. coli* to the C-terminus. As IF1 helps to recruit the 30S subunit of the ribosome to the ribosomal binding site, directing it to a particular mRNA using dCas9 could theoretically allow for increasing mRNA translation. A single-stranded DNA containing a PAM (called a PAMmer) is included in tandem to allow for dCas9 binding to the DNA-RNA region. On the bottom is a depiction of the eight binding sites of sgRNAs designed to test the impact of this system on mCherry fluorescence.

We first designed a two-plasmid system to test dCas9-IF1's effect on the translation rates of mCherry. One plasmid contained either aTc-inducible dCas9, or aTc-inducible dCas9 with IF1 attached to the C terminus. The second plasmid contained the sgRNA sequence that tells the fusion protein where to bind, as well as the mCherry gene under control of the isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible lac promoter. Experiments, were performed in *E.coli* DH5αZ1 expressing both the lac and tet repressors.

9.3.2 dCas9-IF1 modulation of mCherry fluorescence

We first tested the impact of these aTc inducible dCas9-IF1 constructs on IPTG inducible mCherry fluorescence in a DNA-binding context, in the absence of any exogenous PAMmer (Fig. 2). Theoretically, on sgRNAs I1-I6 should be able to allow for dCas9 or dCas9-IF1 binding to the mCherry DNA sequence. Surprisingly, we found significant increases in mCherry fluorescence in the absence of induction of either mCherry or dCas9 induction for the dCas9-IF1 targets I2 and I6, suggesting that mCherry fluorescence was increased by the presence of initiation factor 1 (Fig. 2A). However, we also saw significant (albeit minor) decrease in fluorescence for the dCas9-IF1 targets I4 and I5. This could be due to OD differences in cultures. The reductions were not as large as the I2 increase. With activation of aTc, we saw a significant increase in dCas9-IF1 strain fluorescence using targets I3 and I4, while a significant decrease was observed with I5. This again suggests that initiation factor 1 localization can increase fluorescence, but the mechanism through which this occurs is obtuse at best given the significant decreases of particular strains.

These results are even further complicated by the results observed in the presence of mCherry induction (Fig. 2B). In this case, dCas9-IF1 showed significantly reduced fluorescence with targets I4 and I5 in the absence of dCas9 induction, and I4 and I6 in the presence of dCas9 induction. dCas9-IF1 increased fluorescence slightly with I2, and surprisingly with the non-targeting construct I8.

We attempted to elucidate more clearly the impact of these perturbations by looking at differences in fluorescence with and without induction (Fig 2.C-D). We saw very little differences in dCas9 impacts on fluorescence in the absence of mCherry induction, while we saw very strong increases for targets I3 and I4 in the presence of mCherry induction (Fig. 2C). In the presence of mCherry induction, activation of dCas9 lead to neutral or detrimental impacts on mCherry fluorescence (Fig. 2D). This would make sense, as a context of more mCherry expression should mean more impact from dCas9 binding to the mCherry DNA, blocking transcription. This reduction appeared to be less in dCas9-IF1, especially for targets I4 and I6.

These results do not clearly point to clear activation of gene expression through dCas9-infA binding. However, the increase in gene expression of certain targets, particularly I2, point to the potential of this fusion protein to increase bacterial translation rates. Potentially conflicting factors could arise from the fact that this construct is binding to DNA rather than RNA, and transcription is therefore impacted by dCas9-IF1 binding inside the open reading frame.

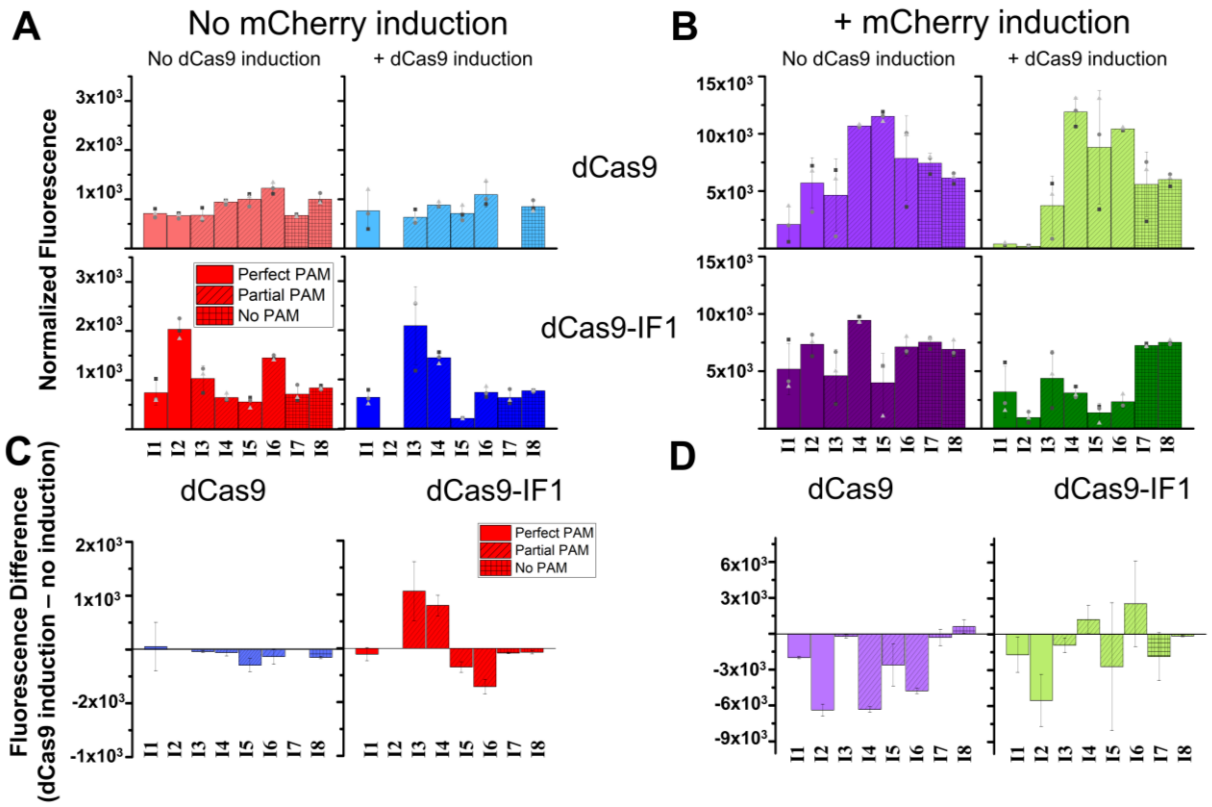


Figure 9.2 Activation of gene expression using dCas9-IF1

Eight constructs were tested for the ability to modulate mCherry fluorescence. Y-axis indicates mCherry fluorescence normalized to the optical density reached after 24 hours of growth under the stated conditions. Fluorescence of strains (A) without IPTG induction or (B) with 1 mM IPTG induction of mCherry fluorescence. Boxes are coded by whether the target contains a perfect NGG PAM, a partially-effective NGA PAM, or an NRR PAM at which dCas9 is completely unable to bind to DNA. The top two graphs show results with dCas9, and the bottom two show results with dCas9-IF1. Stars in the bottom row represent statistically significant differences from the corresponding constructs in the top row ($P < 0.05$). Construct I2 was not tested in dCas9 and dCas9-IF1 with 50 ng/mL aTc induction, nor was construct I7 for dCas9. The differences in the left and right columns were calculated for (C) no mCherry induction and (D) with 1 mM mCherry induction. All error bars represent standard deviation of biological triplicates.

We focused on further clarification of the constructs designed to bind to a perfect PAM site, I1, and I2. We looked at all combinations of dCas9 and mCherry gene activation and the corresponding fluorescence observed (Fig. 3). We saw in this experiment that mCherry fluorescence did not increase significantly when IPTG was added to dCas9 +

target one, but did slightly with dcas9 + target two. However, there were clearly very large increases in fluorescence when dCas9-IF1 constructs had mCherry induced by IPTG addition. Additionally, un-induced mCherry was significantly more fluorescent in dCas9-IF1 constructs targeting site two, as observed in our previous results. These again corroborate the notion that mCherry is somehow being activated by dCas9-IF1 binding to the mCherry DNA. Interestingly though, this occurs optimally when dCas9-IF1 is binding in the middle of the ORF. This could be due to optimal localization to interact with nearby RNAs, or stronger RNAP traffic enabling greater dislocation of dCas9 from the target DNA. While these results are promising, more investigation is needed to clarify exactly how dCas9-IF1 can be used to increase bacterial translation rates.

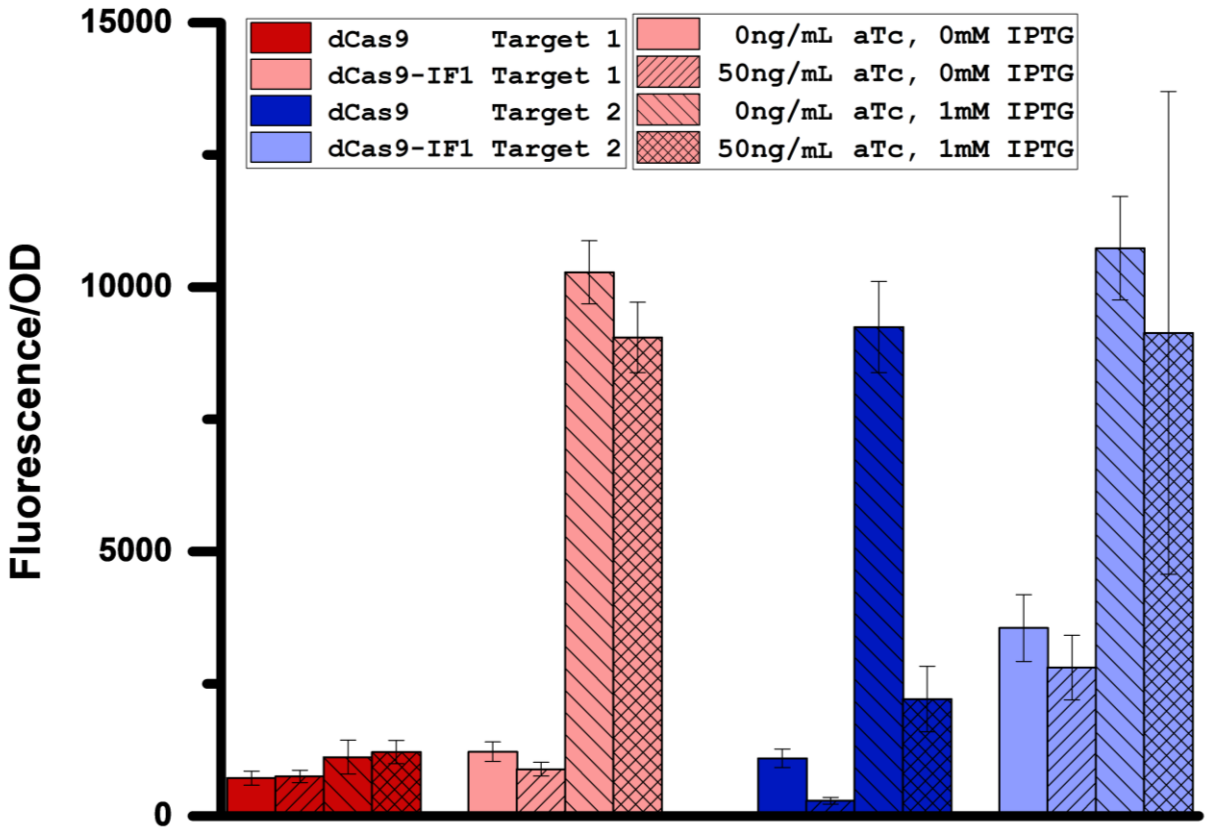


Figure 9.3 mCherry sgRNA targets influence on mCherry fluorescence

Targets I1 (red, left) and I2 (blue, right) were tested for their impact on mCherry fluorescence in the presence of either dCas9 or dCas9-IF1. Crosses in bars represent the state of induction of dCas9/dCas9-IF1 (with aTc) or mCherry (with IPTG). Error bars represent standard deviations of biological triplicates.

9.3.3 Targeting RNA with dCas13a-IF1

During our investigation of utilizing dCas9-IF1 to increase translation rates, a significant advancement in the CRISPR field emerged. Specifically, a novel class of CRISPR enzymes were discovered and shown to utilize a single effector protein to target RNA instead of DNA[70]. This protein, originally called C2c2 and recently renamed Cas13a, presents a more efficient way to target RNA than utilization of the PAMmer system. As with Cas9, Cas13a's ability as a nuclease can be deactivated through two mutations in each of the two Higher Eukaryotes and Prokaryotes Nucleotide-binding Domains (HEPN).

dCas13a can no longer cut sequences, but is able to lock on to specified sequences, as determined by the accompanying CRISPR RNA (crRNA) (Fig. 4A). Through this manner, the issue of DNA-binding effects obfuscating our results is removed using this system.

We therefore decided to revise our experimental design to utilize a deactivated Cas13a fusion to the same initiation factor gene we had previously been utilizing. The design approach is similar in that we utilized dCas13a to lock on to regions near key genes and increase the local concentration of IF1 to drive greater translation rates. The designed crRNAs were engineered taking into consideration g the altered protospacer adjacent motif, which is a preference for a 3' non-G nucleotide protospacer flanking site (PFS)[9], which is only one non-G nucleotide immediately after the target sequence.

Again, we attached IF1 to dCas13a through the EAAKEAAK rigid linker and explored the ability of this fusion protein to modulate the rates of bacterial translation. As we were less sure of which terminus of the protein would be the most effective location to tether IF1, we created both N-terminus and C-terminus fusions. Based on our previous results, the binding location of dCas13a on mRNA could play a significant role in how translation rates are affected. If dCas13a were to bind near the ribosome binding site of the target RNA, it is plausible that ribosomes would be prevented from binding and creating the protein, thus decreasing the rate of translation. We designed a test crRNA targeting the middle of the mCherry ORF sequence (Target 1), as well as three ideal target locations for crRNAs of 62, 80, and 101 nucleotides upstream of the start codon of the mCherry gene (Targets 2-4) to see the impact of these binding sites on gene expression. Predicted mRNA structure of the mCherry gene shows less secondary structure in the area of the Target 1 binding site, while slightly more secondary structure

for Targets 2-4 (Fig. 4B, C)[336]. Through our experiments, we aim to show that the dCas13a-IF1 fusion protein increases the rate of translation in a controlled and tunable manner, through positioning of the fusion protein.

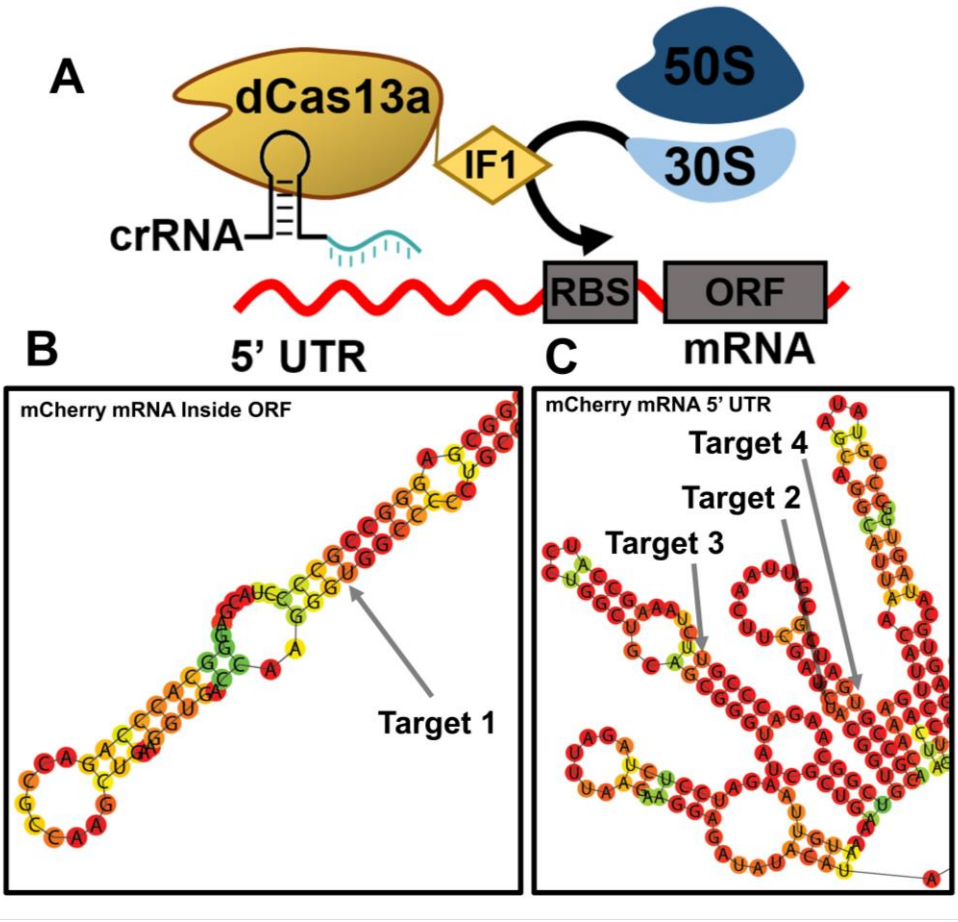


Figure 9.4 Use of dCas13a-infA to increase mRNA translation rates

(A) As before, we designed a fusion protein of dCas13a to the coding sequence of IF1 to increase translation rates. Utilizing a crRNA, this fusion protein can theoretically recruit 30S subunits to mRNA locations, thereby increasing translation rates. (B) Predicted secondary structure of the mCherry Target 1 in the middle of the ORF or (C) the 3 Targets in the 5' UTR. The color scale is based on predicted minimum free energy entropy, with redder colors indicating less entropy.

9.3.4 dCas13a-IF1 driven modulation of mCherry expression

We began by testing how dCas13a and the corresponding IF1 fusions interacted with the crRNA targeting the middle of mCherry ORF to modulate mCherry expression. A nonsense sgRNA target that should unable to complex with dCas13a was used as a control. We should see no significant effect on mCherry fluorescence for conditions of dCas13a/dCas13a-IF1 and the nonsense sgRNA.

As expected, there were no differences between any of the dCas13a variants in the presence of a non-complexing sgRNA (Fig. 5A). Very low fluorescence arising from leaky expression from the lac promoter arose after roughly 40 hours of growth. However, drastic differences in fluorescence were observed between dCas13a and both dCas13a-IF1 variants in the presence of crRNA Target 1 (Fig. 5B). Fluorescence of strains harboring these latter two proteins fluoresced between ~7-8 fold brighter than the strain harboring dCas13a. This clear increase in fluorescence is unlikely to be a result of protein misfolding, as one would expect strains harboring the sgRNA target to fluoresce at these similar levels. These results mirror previous results with dCas9-IF1, wherein the brightest fluorescence was actually achieved by targeting the middle of the mCherry ORF.

We further explored these very promising results by testing how induction with aTc or IPTG affected the cells' fluorescence. Surprisingly, induction of mCherry with IPTG did not result in any significant changes in the fluorescence profiles (Fig. 5 C, D). This could be due to poor localization of the lac promoter on the mRNA leading to poor controllable induction, or degraded IPTG stocks. Induction of these strains with aTc and IPTG resulted in slightly reduced fluorescence of all strains (Fig. 5 E, F), likely due to toxicity of aTc on

the cell. Notably, the C-terminus fusion of dCas13a lost the fluorescence increase previously observed upon induction.

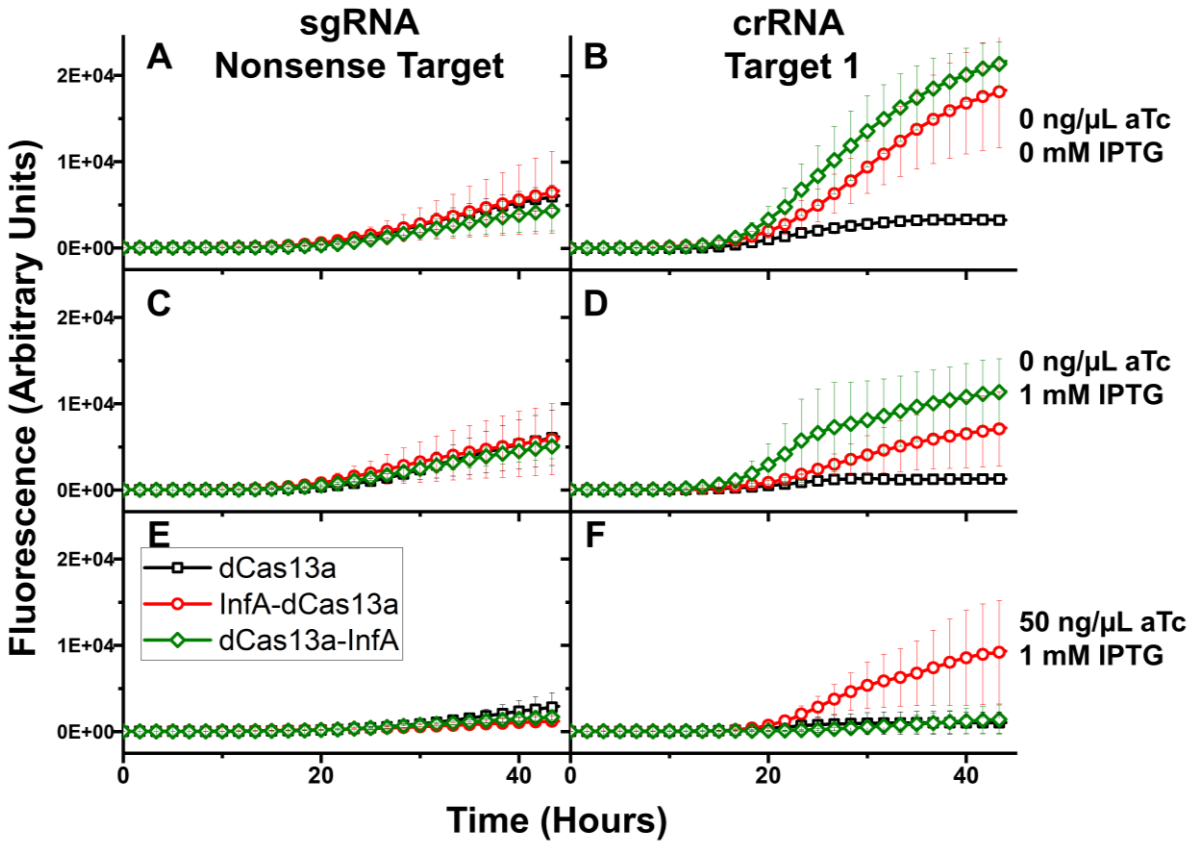


Figure 9.5 dCas13a fused to IF1 increases translation rates

We tested the impact of three different dCas13a variants fused to either nothing, the *infA* gene on the N-terminus (red), or the *infA* gene on the C-terminus (green). The fluorescence of strains harboring these proteins in the presence of either a non-Cas13a complexing sgRNA (left column) or a crRNA targeting the middle of the mCherry ORF (right column) were examined in the presence of either (A,B) no induction, (C,D) induction of mCherry, or (E,F) induction of both mCherry and dCas13a. Error bars represent standard deviations of three biological replicates.

9.3.5 dCas13a-IF1 modulation of mCherry expression at different target locations

In the next experiment, we tested dCas13a and dCas13a-IF1 with several different crRNAs that targeted mCherry at different locations. We focused on just the N-terminus fusion of dCas13a-IF1, as this provided the greatest improvement in fluorescence under

most conditions previously. We targeted dCas13a to both the middle of the mCherry mRNA and several positions upstream of the mCherry mRNA (Fig. 6A). We again tested Target 1, alongside Target 3 which is 60 base pairs (bp) upstream of the translation start site, Target 2 which is 80 bp upstream of the translation start site, and Target 4 which is 100 bp upstream of the translation start site. Finally, we tested a control Target 5, which directs dCas13a/dCas13a-IF1 to bind to green fluorescent protein (GFP) which is not present in the cell. Target 5 acts as another control in that dCas13a/dCas13a-IF1 with this crRNA should not be able to target any specific RNA in the cell. We also retested the nonsense sgRNA as another control. Note that the only difference between Target 005 and the nonsense sgRNA control is that Target 005 should be processed by dCas13a, such that the pre-crRNA is modified to crRNA, and is then able to patrol the cell with dCas13a to look for a match, however neither should be able to bind to any target RNA in the cell.

Each crRNA was transformed into a DH5αZ1 strain of *E.coli* with either dCas13a or dCas13a-IF1, and mCherry fluorescence was measured over time. For Target 1, we found similar findings compared to the last experiment, where the cells containing dCas13a-IF1 had significantly greater mCherry fluorescence than cells containing dCas13a alone (Fig. 6B). However, we began to see significant similarities between dcas13a and dCas13a in all of the other target crRNAs (Fig. 6C-E). And while the nonsense sgRNA behaved similarly to the previous experiment in that very low expression was seen for both dCas13a and dCas13a-IF1, the nonsense crRNA target surprisingly showed significantly higher mCherry fluorescence (Fig. 6F-G).

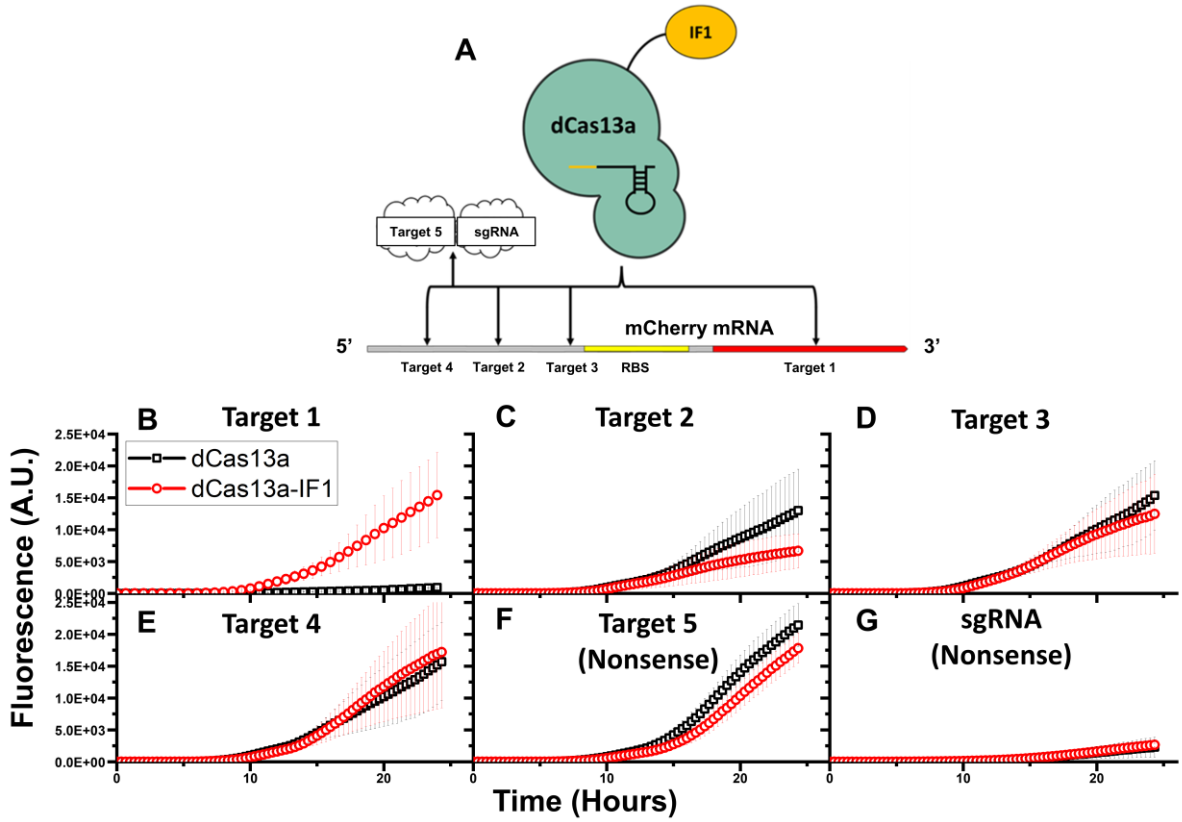


Figure 9.6 Target location influence on mCherry fluorescence

(A) We tested the various different crRNA targets for their ability to influence mCherry fluorescence. Target 1 from the previous figure was compared against three targets upstream of the RBS, as well as the previous control sgRNA and a new GFP-targeting crRNA. The fluorescence profiles in the absence of induction of either IPTG or aTc are presented in B-G for the various different crRNA or sgRNA guides. Error bars represent standard deviation of biological triplicates.

9.3.6 dCas13a-IF1 modulation of genome-based lacZ expression

While performing experiments to determine the effect dCas13a-IF1 might have on translation rates of mCherry, we also wanted to explore if the fusion protein could increase translation rates for a gene that was naturally found in *E.coli*. We chose the lacZ gene encoding β-galactosidase, which is a part of the lac operon that allows *E.coli* to metabolize lactose. In this case, we can measure changes in translation rates using the Miller Assay, a fluorescent assay that measures the concentration of β-galactosidase. In

the assay, samples are given an excess of a synthetic compound called o-nitrophenyl-β-D-galactoside (ONPG) which is then cleaved by β-galactosidase to produce o-nitrophenol, which has a yellow color. Therefore, samples that have a greater expression of lacZ will glow yellow at a greater intensity than other samples.

For this experiment, we designed two additional crRNAs that targeted upstream of the lacZ gene. The first target, crRNA lacZ1, binds in an area slightly overlapping the ribosome binding site for the lacZ RNA, while the second target, crRNA lacZ 2, binds within the lacZ RNA (Fig. 7A). We also tested the nonsense sgRNA that we had used for the mCherry experiments. We tested all of the crRNAs and the nonsense sgRNA with dCas13a and N-terminus dCas13a-IF1, as well as a control of the DH5αZ1 strain. We also treated all conditions with either 0 ng/mL aTc, or 10 ng/mL aTc, to see if overexpression of dCas13a/dCas13a-IF1 had any effect on expression of lacZ.

Miller units for strains under each of these conditions are reported in Fig. 7B, corresponding to the level of lacZ expression. We observed the highest level of lacZ expression in *E. coli* harboring no plasmids, likely due to slightly faster growth because of the lack of plasmid burden. Both dCas13a and dCas13a-IF1 exhibited slightly higher lacZ expression utilizing the 2nd target over the 1st target, with very little differences between either dCas13a or dCas13a-IF1 in either condition. However, this trend changed slightly upon induction with aTc. While lacZ expression decreased in both dCas13a strains as well as the empty *E. coli* strain, expression was maintained or increased in the dCas13a-IF1 strains. This is best demonstrated by comparing the differences with and without induction for each strain, as shown in Fig. 7C.

The finding that lacZ1 targets exhibited lower expression than corresponding lacZ2 targets is likely due to where the fusion proteins are binding to. The crRNA lacZ1 directs dCas13a/dCAs13a-IF1 to bind in a region that overlaps with the ribosome binding site for the lacZ RNA, while crRNA lacZ2 directs dCas13a/dCas13a-IF1 to bind slightly downstream of the translation start site. Both of these positions should have a negative effect on translation because they act as roadblocks for the ribosomes translating the RNA. This effect is likely more pronounced for crRNA lacZ1 because it overlaps the ribosome binding site itself, as seen in the significant drop in Miller Units for dCas13a with crRNA lacZ1 when aTc is added. While these results do suggest that there is potentially a beneficial effect from the presence of the fusion protein on rates of translation, more work is needed to fully characterize how IF1 fusion proteins can be utilized to increase translation rates.

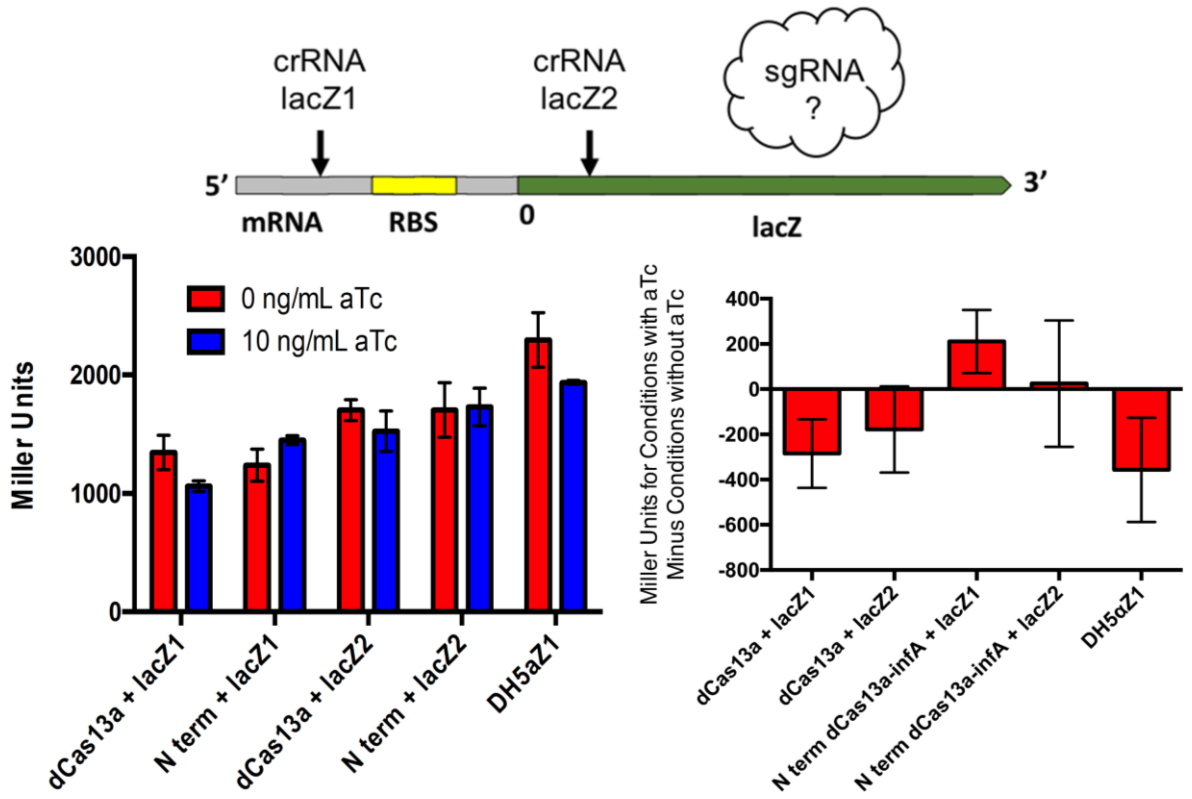


Figure 9.7 Miller Assay investigating dCas13a-IF1 impact on *lacZ* expression

(A) A map showing the location of where crRNAs target the mRNA of lacZ. There are four nucleotides in which the lacZ1 target overlaps the RBS, while lacZ2 targets 220 nucleotides downstream of the translation start site. (B) Graph showing the Miller Units calculated for five different conditions as noted. Each condition was treated with either no aTc (red bars) or 10 ng/mL aTc (blue bars). Error bars represent standard deviation of four biological replicates. (C) The difference between the red and blue bars of (B) was calculated to show the impact of aTc induction on lacZ expression. Error bars represent corresponding standard deviations calculated using the appropriate error propagation formula.

9.4 Discussion

In this study, we designed novel fusion proteins of dCas9 and dCas13a with IF1 for modulating the translation rates of specific target RNAs. We have collected promising results showing significant increases in gene expression between these fusion proteins and their base, non-fusion proteins (see Figs. 3 and 5). The increased expression we observe appear to be due to modifications of translation rates. However, the mechanism of how this occurs remains quite unclear.

Gene expression increase was first observed in the context of dCas9 binding in the middle of the mCherry ORF. One would predict decreases in gene expression due to transcriptional interference, so the fact that these targets lead to increases in gene expression is surprising. However, our results also show very clearly that this approach was not universally successful. Indeed, the potential for transcriptional interference and the difficulty in delivering PAMmers indicates to us that this approach will likely be the less viable pathway to proceed with.

However, these results do suggest key influences underlying how future fusion proteins should be designed. Most notable is the potential impact of molecule stoichiometry. It should be highlighted that the greatest success we observed was through leaky expression of both dCas9-IF1 and mCherry, suggesting that greater gene expression increase can be achieved by sparsely targeting RNAs with low basal concentrations. It is highly likely that the concentration of IF1, both from the genome as well as from the fusion protein, has a major role as well. *In vitro* translation experiments in which each of these molecules can be varied independently will help to elucidate specific rules of optimal stoichiometry.

Additionally, rules for where to target the CRISPR-IF1 fusion proteins need to be more clearly defined. The fact that the best success we observed was by targeting inside the ORF of the target genes suggests that these fusion proteins may not be actually influencing the specific mRNAs they are targeting. Rather, it may be that the nearby RNAs that the fusion protein binds to are being influenced. As translation in bacteria usually takes place in the immediate vicinity of where the RNA is produced, this could explain why targeting the DNA can result in increased translation rates. Similarly, by “sacrificing” one RNA who is targeted in the middle of its ORF, these fusion proteins could potentially still increase translation rates of the nearby RNAs. In a similar vein, this could explain why all of the crRNAs guiding dCas13a showed very similar results for the fusion proteins. Even the non-targeting crRNA Target 5, since it is expressed on the same plasmid as the mCherry mRNA, could draw in dCas13a-IF1 to the local area of the cell where that RNA is being expressed. Such collateral effect would again be best explored through *in vitro* assays to minimize confounding variables.

Future work into creating a dCas13a-IF1 fusion system for increasing translation rates would also benefit from studies further characterizing IF1, and specifically how mutations of the translation factor could be harnessed. For instance, it might be beneficial to abolish certain functions of the initiation factor, such as its ability to complex with fully assembled ribosomes to prevent potential off-targeting effect. These modifications could be done through additional cloning or through protein engineering, using either directed evolution or rational protein design. Additionally, we could look into the feasibility of using a different initiation factor or a different protein altogether that could, either through ribosome recruitment or a different mechanism, increase translation rates.

9.5 Materials

9.5.1 Bacterial Strains, Media, and Culture Conditions

E.coli cloning strains NEB 10-β (New England Biolabs) and the final experimental strain DH5αZ1 (Zymo Research Corporation) were cultured in Lysogen Broth (LB) (Sigma-Aldrich), except for experiments with mCherry incorporated into the *E.coli* genome, which used M9 minimal media (5X M9 minimal media salts solution from MP Biomedicals, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$ in sterile water) with 0.4% weight/vol glucose (34.2m mM). Colonies were grown on LB-agar plates and, depending on the plasmid, either received no antibiotic, chloramphenicol only (35 μg/mL), or both chloramphenicol (35 μg/mL) and ampicillin (100 μg/mL). Cultures (5 mL) were grown at 37℃ with constant shaking at 225 r.p.m. All experiments used biological triplicates, unless otherwise noted, which were inoculated from individual colonies grown on LB-agar supplemented with the appropriate antibiotic.

9.5.2 dCas9-infA construction

dCas9 was obtained from Addgene Plasmid 44249 from the Qi lab. The coding sequence of *infA* was amplified from the *E. coli* genome using the primers GAAGCTGCCGCAAAGGAAGCTGCGGCCAAGGCCAAAGAAGACAATATTGAAATGCAAGGT and TTATTTGATGCCTGGAGATCCTTACTCGAGTTAGCGACTACGGAAGACAATGCGGCCTTT, which was then attached to the dCas9 plasmid using the Gibson assembly primers AAAGGCCGCATTGTCTTCCGTAGTCGCTAACTCGAGTAAGGATCTCCAGGCATCAAATAA and AAAGGCCGCATTGTCTTCCGTAGTCGCTAACTCGAGTAAGGATCTCCAGGCATCAAATAA.

9.5.3 dCas13a-infA construction

dCas13a was obtained from Addgene Plasmid 83485 from the Doudna lab. The coding sequence of dCas13a was first cloned into the same vector as a previously designed vector, dCas9 (Addgene Plasmid 44249), under the same aTc inducible promotor, rrnB T1 terminator, and chloramphenicol resistance marker. PCR with Phusion High-Fidelity DNA Polymerase (New England Biolabs) was used to amplify these both the existing dCas9 vector and the dCas13a vector. These were subsequently PCR purified (GeneJET PCR Purification Kit, Thermo Scientific) and digested with *ApaI* and *XhoI* (Restriction Endonuclease, New England Biolabs) as per the provided protocols. Both the insert and backbone were gel purified (Zymoclean Gel DNA Recovery Kit, Zymo Research Corporation). The backbone was additionally treated with Alkaline Phosphatase and PCR purified. The insert and backbone were ligated together using T4 DNA Ligase (Thermo Scientific), PCR purified, and transformed into electrocompetent NEB 10-β cells. Plasmids were miniprepped using Zyppy Plasmid Miniprep Kit (Zymo Research Corporation). Sequencing of final sgRNA constructs was performed for validation of correct assembly product (GENEWIZ).

To construct the fusion protein, oligonucleotide sequences were ordered that contained the gene for initiation factor 1 (*infA*) for both the N term and C term orientation. PCR was then performed to amplify the insert (*infA*) and backbone (dCas13a), and subsequently gel purified. The plasmid was assembled using Gibson Assembly (New England Biosciences).

9.5.4 Guide RNA target design

The different crRNA and sgRNA plasmids for the two-plasmid mCherry experiments were assembled in a similar fashion as described above, except that primers were used to replace the crRNA target sequence of one previously existing crRNA, with the new target crRNA sequence. Additionally, both the backbone and insert were digested with *BglII* and *BamHI* (Restriction Endonuclease, New England Biolabs).

The crRNA plasmids for the mCherry experiments were then modified for the experiments where mCherry was integrated into the *E.coli* genome. The specific crRNA sequence was cut out of the plasmid containing mCherry and the crRNA sequence using *AatII* and *BamHI-HF* (Restriction Endonuclease, New England Biolabs). A previously designed plasmid that did not contain the mCherry gene was also digested with these two restriction endonucleases.

To construct the crRNA plasmids for the lacZ system, the crRNA sequence from one of the crRNA plasmids used for the mCherry experiments was replaced with crRNA sequences targeting the lacZ RNA using primers. The backbone and insert were then digested with *HindIII-HF* and *AatII* (Restriction Endonuclease, New England Biolabs).

9.5.5 Microplate experiments for mCherry fluorescence readings

Biological triplicates, unless otherwise stated, were inoculated from individual colonies of DH5αZ1 cultures harboring either dCas13a/dCas13a-IF1 and crRNA plasmids into 5 mL LB cultures with ampicillin and chloramphenicol. These were grown overnight for 16 hours to stationary phase. The next day, 2 µL of overnight culture was used to

inoculate 198 µL of LB with antibiotics in a 96-well microplate. Depending on the experiment, conditions were also induced with aTc, IPTG, or both at varying levels (specified in the Results section above). A Tecan GENios microplate was used to track fluorescence or optical density over time with continuous shaking. Optical densities were measured at 590 nm absorbance and mCherry fluorescence was measured at 620 nm in 20 min intervals. Temperature was maintained at 37℃, and cultures were shaken for 16.6 min after each measurement with an additional 10 s of shaking before measurement.

9.5.6 Miller Assay protocol

Cultures were grown overnight for 16 hours in 5 mL LB supplemented with chloramphenicol and ampicillin. In the morning, cultures were diluted 1:100 in 2 mL LB cultures, supplemented with IPTG (1mM), and allowed to grow for 2-3 hours until they reached mid-log phase (OD 0.4-0.8). After this time, 200 µL of culture was transferred to a 96-well plate and absorbance at 600 nm was measured. 20 µL of the culture was then removed and added to 80 µL of permeabilization solution (100 mM Na_2HPO_4, 20 mM KCl, and 5.4 µL/mL β-mercaptoethanol in sterile water), 10 µL of chloroform, and 5 µL of 0.1% SDS. Samples were then warmed at 30℃ for 20-30 minutes. Afterward, 600 µL of substrate solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 1 mg/mL ONPG, and 2.7 µL/mL β-mercaptoethanol in sterile water) was added and the sample was vortexed. Additionally, the exact time of addition of the substrate solution was recorded down to the second. The sample was then placed on the hot block and observed until it reached a medium yellow color. When this occurred, 700 µL of stop solution (4.75E-5 M Na_2CO_3) was added and the sample was vortexed. Additionally, the exact time of addition of the substrate solution

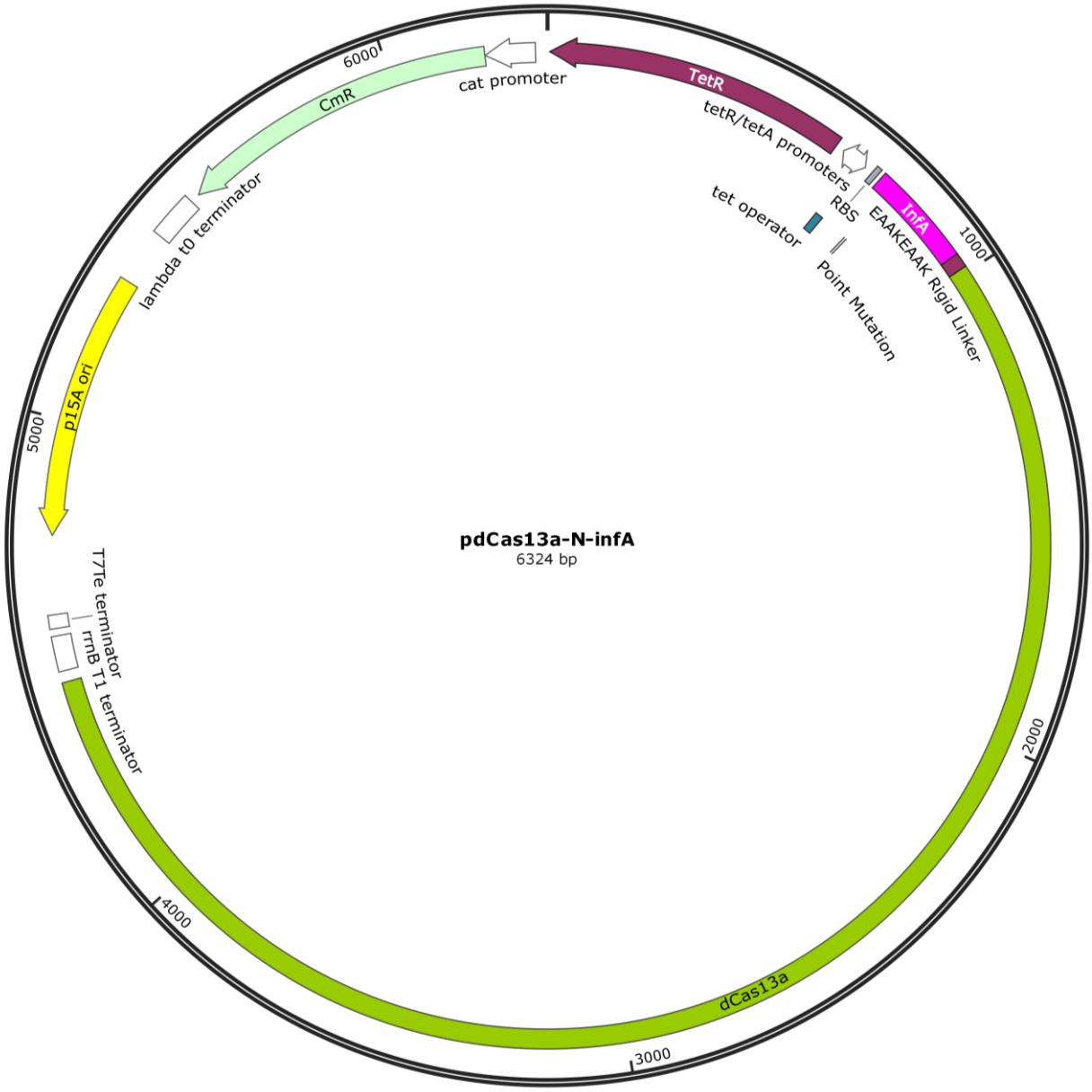
was recorded down to the second. If this color change did not occur within 2 hours, the stop solution was added anyway. The samples were then centrifuged at 15,000 RPM for 10 minutes. After centrifugation, 200 µL of the supernatant was added to a 96-well plate and the absorbance at 420 nm and 550 nm was recorded. The Miller Units for each sample was then calculated using the following equation: $1\,Miller\,Unit = 1000 * \frac{Abs_{420} - (1.75 * Abs_{550})}{t * v * Abs_{600}}$, where Abs_{420} is the absorbance of the yellow o-nitrophenol, Abs_{550} is the scatter from the cell debris, t is the reaction time in minutes, v is the volume of culture assayed in milliliters, and Abs_{600} reflects cell density.

This is calculated from absorbance measurements before and after the assay and the time it took for the assay to be completed. Samples with higher Miller Units indicate that lacZ had a greater rate of expression, causing more β-galactosidase to be made, which could then cleave more o-nitrophenyl-β-D-galactoside, and thus produce more yellow-colored o-nitrophenol.

9.6 Author Contributions

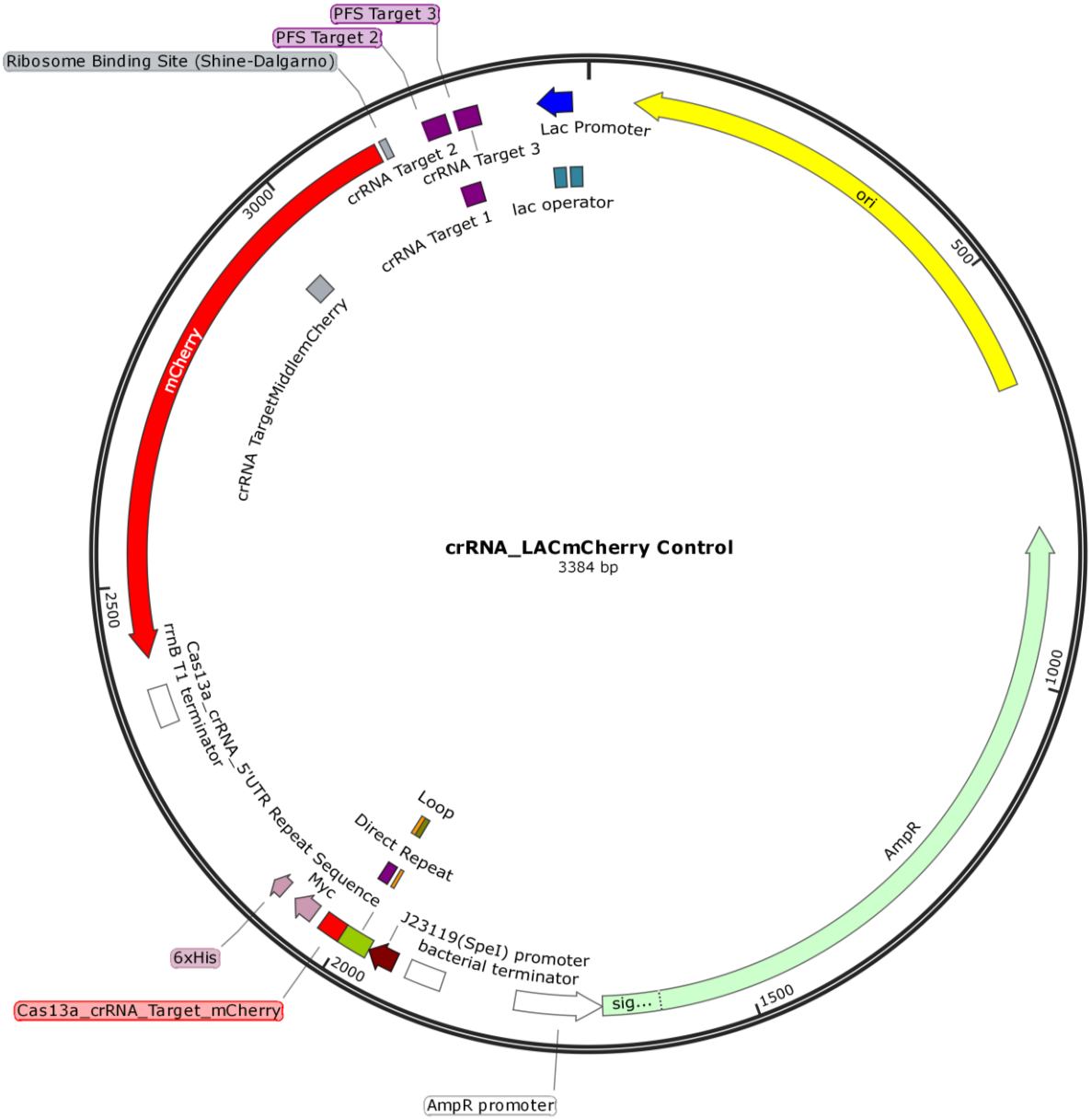
P.B.O., M.J.S., and A.C. devised and planned the study. M.J.S. and P.B.O performed all experiments and wrote the manuscript. M.J.S., W.T.C., and P.B.O. constructed all plasmids. P.B.O. and A.C. revised the manuscript.

9.7 Supplementary Info



Supplementary Figure 9.1 Plasmid map of dCas13a-IF1

The N-terminus fusion of dCas13a and IF1 is shown here. The same map was used for all other dCas9 and dCas13a constructs where the area between the Tet promoter and rrnB terminator are replaced with the corresponding sequences. Note that a point mutation (resulting in no change in amino acid sequence) was made in *infA* in order to allow for cloning with KpnI.



Supplementary Figure 9.2 Plasmid map of crRNA

An example of the crRNA/sgRNA plasmids harboring mCherry is shown here. The same plasmid map is used for all other crRNAs, where the short 28 nt “Cas13a_crRNA_target_mCherry” sequence is replaced to encode for the target sequence. This sequence, as well as the upstream stem-loop structure, is replaced with the corresponding sgRNA sequences as in Addgene plasmid 44251.

Chapter 10: Concluding Remarks

10.1 Summary

In my thesis, I have demonstrated how gene expression engineering is a viable alternative to traditional genome engineering techniques for modifying bacterial systems. Particularly, I have utilized *a posteriori* knowledge (Chapter 3) to engineering improved bacterial biofuel tolerance (Chapter 4) or reduced bacterial antibiotic tolerance (Chapter 5). I have extrapolated results learned in these chapters to engineering epistatic (Chapter 6) and synergistic (Chapter 7) interactions for controlling adaptive trajectories, or resensitizing antibiotic resistant bacteria. Finally, I have outlined novel CRISPR techniques for engineering gene expression (Chapters 8 and 9).

How bacteria naturally respond to biofuel and antibiotic stressors were first identified in Chapter 3, yielding a set of genes whose expression served as promising candidates for manipulating in order to impart desired phenotypes. A number of unique approaches were taken to identify these gene candidates. This includes the identification of how gene expression stochasticity differed in both the presence and absence of antimicrobial stress. Changes in the variability of gene expression have been a here-to-for unappreciated aspect of gene expression that we argue are critically underexplored. Increases in variation of particular genes could indicate that bacteria are exploring wide variety of states, and such genes might be more important for manipulating how bacteria adapt to antimicrobial stress conditions.

Chapters 4-7 represent the bulk of this thesis, in which knowledge gained from Chapter 3 was applied to rationally engineer bacterial response to stress. I began by showing how 31 CRISPR interference and CRISPR activation perturbations of *E. coli*

caused significant modifications in biofuel tolerance (Chapter 4). This revealed a number of promising and here-to-fore unexplored approaches to improve *E. coli* n-hexane and n-butanol tolerance, such as repressing *wzc* or *yjjZ* expression. I next explored the converse of this, in which CRISPR perturbations were applied to reduce bacterial fitness during antibiotic exposure (Chapter 5). This enabled reversible control over fitness during the early stages of treatment, demonstrating how modifying gene expression could be applied to influence bacterial response to antimicrobials. Importantly, I also identified a unique and surprising phenomenon in which combining gene perturbations resulted in worse growth than was expected. Multiplexing appeared to reduce fitness in a fashion greater than the sum of its parts.

This discovery catalyzed a thorough exploration into how simultaneous gene expression changes interacted with one another to influence bacterial fitness. In Chapter 6, I systematically merged CRISPR inhibition and activation constructs to manipulate expression of one to four genes at a time, and calculated the influence such perturbations had on one another upon combination. This revealed a stunning trend towards nonlinear reductions in fitness that was first observed in Chapter 5; the more perturbations that were combined, the more fitness losses were exacerbated. Indeed, these fitness losses were greater than the sum of their parts, and in some instances even showed that combining positive fitness perturbations reduced fitness even more than was to be expected. This bears striking similarity to the phenomenon known as epistasis, which is traditionally used to describe the non-linear phenotypes that arise when combining two or more mutations. Only a handful of studies have even acknowledged that such a phenomenon might arise at the gene expression level. Our study was the first demonstration that it is possible to

actually artificially introduce negative epistasis by multiplexing perturbations of gene expression.

Furthermore, epistasis has a critical role in evolutionary theory. In particular, negative (or positive) epistasis has been described as a fundamental force that restricts (or promotes) particular adaptive trajectories available to an organism. Epistasis therefore acts as a mechanism controlling how organisms evolve. I hypothesized that if our multiplexed perturbations were systematically introducing negative epistasis, then this could be exploited to control how bacteria adapt. In the latter parts of Chapter 6, I demonstrate that indeed, strains harboring multiplexed perturbations adapted slower to antibiotics during the initial stages of treatment. Over a longer-term of treatment typical of clinical antibiotic prescriptions, CRISPR multiplexed perturbations prevented *E. coli* from reaching similar levels of antibiotic tolerance as a control strain, even when the experiment was explicitly biased by propagating only the most tolerant strains of the multiplexed strain. Overall, Chapter 6 suggests that it may be possible to design antibiotics that directly combat the evolution of antibiotic resistance.

I built upon the concepts of harnessing genetic interactions for antimicrobial purposes developed in Chapter 6 in our next study (Chapter 7). In this case, I explored the use of gene expression perturbations to potentiate antimicrobials, and even resensitize multidrug-resistant strains to antibiotic treatment. This involved harnessing the power of synergy, a therapeutic concept highly analogous to epistasis. While synergy in a treatment context typically involves combining two or more antibiotics to reduce fitness in a non-linear fashion, here this concept is applied in a novel fashion. Specifically, we focused on developing gene expression perturbation treatments that pose no inherent

fitness cost individually to potentiate another antibiotic. Through a systematic exploration of 270 gene-drug knockouts, we developed CRISPRi and PNA therapies to replicate antibiotic synergy using a gene knockdown context. These therapies were shown to facilitate curing of intracellular Salmonella infections, as well as resensitizing clinically isolated multi-drug resistant bacteria to antibiotic treatment. This paves the way for a novel class of antibiotic adjuvants in the perpetual fight against antibiotic resistance.

Finally, I describe novel CRISPR gene expression engineering strategies that we have developed in Chapters 8 and 9. These techniques, based on the recently discovered Cas13a RNA-targeting CRISPR system, a “smart” antibiotic to degrade antibiotic resistance genes’ mRNA and a fusion of Cas13a to initiation factor 1 to control bacterial RNA translation rates. Furthermore, in Chapter 4, I outline the use of a hyper-mutator system to rapidly generate new sgRNA targets. While each of these technologies is still in their infancies, they each represent promising avenues of future research.

Taken together, this thesis demonstrates our capacity to exploit the native biological processes that lead to inherent heterogeneity in gene expression to engineer bacteria for desired purposes.

10.2 Gene Expression Engineering: An Ethical Argument

It would be frivolous to say that life is complicated. The intricate biological processes that comprise the fundamental building blocks of organisms have been honed across millennia, resulting in systems that are profound, convoluted, and virtually inscrutable. Nevertheless, the era of synthetic biology is upon us. Novel tools to engineer life are

rapidly emerging. A zealous and infectious enthusiasm for the field's potential to revolutionize humanity is taking hold, even among the layman.

As the world races to embrace the possibilities these technologies offer, serious and difficult discussions regarding the ethical implications are emerging. This has primarily been driven by two demonstrations from China of Cas9 editing in human germ line cells, the first of which was rejected from both Science and Nature on largely ethical concerns[337–339]. Scientists across the globe have rallied to point out significant pitfalls of our current genome engineering techniques, such as the potential of off-targets that inadvertently introduce problematic mutations[340–342]. Furthermore, cultural considerations such as the ability to enact ideals of Social Darwinism raise the potential for drastic detrimental societal implications. While alternative viewpoints certainly exist, the consensus viewpoint is best described in a prominent Nature article arguing that the editing of the human germline "is dangerous and ethically unacceptable" at present[343]. The overwhelming power these new synthetic biology tools have given humanity is, thankfully, being met with overwhelming caution. However, there is no explicit law in the United States preventing genome engineering of human germline cells, and researchers in the US have successfully demonstrated successful editing of a human embryo[344].

It would be naïve to assume that these significant concerns have no bearing outside the context of the human germline. As the ethical debate moves into the limelight, its ramifications will likely extend to all applications of genome editing. The potential for conflation of genome editing in somatic versus germline cells in the public debate is conceivable, if not highly probable without educated leaders directing the discussion. Regulation through agencies such as the US Department of Agriculture will likely be

influenced by this debate, which could result in restrictions or outright bands in genome engineering applications in all human cells. This would increase the barrier in applying technologies such as CRISPR in combating antibiotic resistant infections in clinical settings.

Regardless of whether these possibilities come to fruition, it behooves the scientific community to continue investigating alternative approaches to genetic engineering that achieve similar desired outcomes. This is where the ethical benefits of gene expression engineering clearly surpass other techniques which directly alter the genome. Techniques such as dCas9, Cas13, and PNA used in this thesis do not exhibit the potential to enact direct, heritable changes to a genome. Still, they can be applied to alter how the genome is read in an organism. Embracing gene expression engineering can avoid many of the complicated and difficult philosophical and ethical questions that accompany DNA modifying genome engineering techniques.

In this thesis, I have focused primarily on the application of gene expression engineering towards increasing biofuel tolerance or interfering with bacterial tolerance to antimicrobials. However, the potential for these applications towards higher-order organisms is momentous, and still in its infancy. Additionally, there are reasons to believe that gene expression engineering will be more applicable to eukaryotic systems, especially through techniques that directly target RNA such as PNA or Cas13a. Since these RNA engineering approaches do not require entering the nucleus, they can be applied with significantly less technical or ethical concern for interfering with or modifying heritable DNA elements.

Chapter 11: Bibliography

1. Elowitz, M. B. & Leibier, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338. DOI: 10.1038/35002125.
2. Khalil, A. S. & Collins, J. J. (2010) Synthetic biology: Applications come of age. *Nat. Rev. Genet.* **11,** 367–379. DOI: 10.1038/nrg2775.
3. Purnick, P. E. M. & Weiss, R. (2009) The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.* **10,** 410–422. DOI: 10.1038/nrm2698.
4. Keasling, J. D. (1999) Gene-expression tools for the metabolic engineering of bacteria. *Trends Biotechnol.* **17,** 452–460. DOI: 10.1016/S0167-7799(99)01376-1.
5. Niu, W., Guo, J. & Van Dien, S. (2016) *Heterologous pathway engineering*. *Metabolic Engineering for Bioprocess Commercialization* DOI: 10.1007/978-3-319-41966-4_3.
6. Song, C. W., Lee, J. & Lee, S. Y. (2015) Genome engineering and gene expression control for bacterial strain development. *Biotechnol. J.* **10,** 56–68. DOI: 10.1002/biot.201400057.
7. Blazeck, J. *et al.* (2014) Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat. Commun.* **5,** 3131. DOI: 10.1038/ncomms4131.
8. Liu, L. *et al.* (2015) Surveying the lipogenesis landscape in *Yarrowia lipolytica* through understanding the function of a Mga2p regulatory protein mutant. *Metab. Eng.* 1–10. DOI: 10.1016/j.ymben.2015.07.004.
9. Ragauskas, A. J. *et al.* (2006) The path forward for biofuels and biomaterials. *Science* **311,** 484–489. DOI: 10.1126/science.1114736.
10. Bugg, T. D. H., Ahmad, M., Hardiman, E. M. & Singh, R. (2011) The emerging role for bacteria in lignin degradation and bio-product formation. *Curr. Opin. Biotechnol.* **22,** 394–400. DOI: 10.1016/j.copbio.2010.10.009.
11. Agler, M. T., Wrenn, B. A., Zinder, S. H. & Angenent, L. T. (2011) Waste to bioproduct conversion with undefined mixed cultures: The carboxylate platform. *Trends Biotechnol.* **29,** 70–78. DOI: 10.1016/j.tibtech.2010.11.006.
12. Quandt, E. M. *et al.* (2013) Decaffeination and measurement of caffeine content by addicted escherichia coli with a refactored N-demethylation operon from pseudomonas putida CBB5. *ACS Synth. Biol.* **2,** 301–307. DOI: 10.1021/sb4000146.
13. Ventola, C. L. (2015) The antibiotic resistance crisis: part 1: causes and threats. *Pharm. Ther.* **40,** 277–83.
14. Center for Disease Control and Prevention. Antibiotic resistance threats in the United States. Available at www.cdc.gov/drugresistance/threat-report-2013/. Accessed February 5, 2018. (2013). DOI: CS239559-B.
15. Spellberg, B. *et al.* (2008) The epidemic of antibiotic-resistant infections: a call to

action for the medical community from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **46,** 155–64. DOI: 10.1086/524891.

16. Erickson, K. E., Otoupal, P. B. & Chatterjee, A. (2015) Gene Expression Variability Underlies Adaptive Resistance in Phenotypically Heterogeneous Bacterial Populations. *ACS Infect. Dis.* 555–567. DOI: 10.1021/acsinfecdis.5b00095.

17. Erickson, K. E., Otoupal, P. B. & Chatterjee, A. (2017) Transcriptome-Level Signatures in Gene Expression and Gene Expression Variability during Bacterial Adaptive Evolution. *mSphere* **2,** 1–17. DOI: 10.1128/mSphere.00009-17.

18. Otoupal, P. B. & Chatterjee, A. (2018) CRISPR gene perturbations provide insights for improving bacterial biofuel tolerance. *Front. Bioeng. Biotechnol.* In Review.

19. Otoupal, P. B., Erickson, K. E., Bordoy, A. E. & Chatterjee, A. (2017) CRISPR perturbation of gene expression alters bacterial fitness under stress and reveals underlying epistatic constraints. *ACS Synth. Biol.* **6,** 94–107. DOI: 10.1021/acssynbio.6b00050.

20. Otoupal, P. B., Cordell, W. T., Bachu, V., Sitton, M. J. & Chatterjee, A. (2018) CHAOS: Deterring Bacterial Adapation via Epistatic Gene Expression Perturbations. *Nat. Commun. Biol.* In Review.

21. Otoupal, P. B., Erickson, K. E., Eller, K. A. & Chatterjee, A. Conditional killing: engineering fitness-neutral gene perturbations to resensitize multidrug resistant bacteria to treatment. *In Preperation*

22. Otoupal, P. B., Cordell, W. T., Sitton, M. J. & Chatterjee, A. Next generation “smart” antibiotics using a Holin-Cas13a kill switch. *In Preperation*

23. Otoupal, P. B., Sitton, M. J., Cordell, W. T., Bachu, V. & Chatterjee, A. Engineering sequence-specific control over bacterial translation rates using deactivated Cas13a fused to IF1. *In Preperation*

24. Crick, F. (1970) Central dogma of molecular biology. *Nature* **227,** 561–563. DOI: 10.1038/227561a0.

25. Voet, D. & Voet, J. G. (2011) *Biochemistry*.

26. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Pnas* **102,** 12678–12683.

27. Pfleger, B. F., Pitera, D. J., Smolke, C. D. & Keasling, J. D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24,** 1027–32. DOI: 10.1038/nbt1226.

28. Wang, H. H. *et al.* (2012) Genome-scale promoter engineering by coselection MAGE. **9,** DOI: 10.1038/nmeth.1971.

29. Kabadi, A. M. & Gersbach, C. a (2014) Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods* **69,** 188–197. DOI: 10.1016/j.ymeth.2014.06.014.

30. Liu, L., Otoupal, P., Pan, A. & Alper, H. S. (2014) Increasing expression level and copy number of a Yarrowia lipolytica plasmid through regulated centromere

function. *FEMS Yeast Res.* DOI: 10.1111/1567-1364.12201.

31. Hille, F. *et al.* (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172,** 1239–1259. DOI: 10.1016/j.cell.2017.11.032.

32. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169,** 5429–33.

33. Mojica, F. J. M., Juez, G. & Rodriguez-Valera, F. (1993) Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* **9,** 613–621. DOI: 10.1111/j.1365-2958.1993.tb01721.x.

34. Lander, E. S. (2016) The Heroes of CRISPR. *Cell* **164,** 18–28. DOI: 10.1016/j.cell.2015.12.041.

35. Mojica, F. J. M. & Rodriguez-Valera, F. (2016) The discovery of CRISPR in archaea and bacteria. *FEBS J.* **283,** 3162–3169. DOI: 10.1111/febs.13766.

36. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182. DOI: 10.1007/s00239-004-0046-3.

37. Barrangou, R. *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–12. DOI: 10.1126/science.1138140.

38. Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821. DOI: 10.1126/science.1225829.

39. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–9. DOI: 10.1038/nbt.2508.

40. Ledford, H. (2015) CRISPR, the disruptor. *Nature* **522,** 20–24. DOI: 10.1038/522020a.

41. Ledford, H. (2016) Riding the CRISPR wave. *Nature* **531,** 156–159. DOI: 10.1038/531156a.

42. Hwang, W. Y. *et al.* (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31,** 227–229. DOI: 10.1038/nbt.2501.

43. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–23. DOI: 10.1126/science.1231143.

44. Qi, L. S. *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–83. DOI: 10.1016/j.cell.2013.02.022.

45. Bikard, D. *et al.* (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41,** 7429–7437.

46. Abudayyeh, O. O. *et al.* (2017) RNA targeting with CRISPR–Cas13. *Nature* **550,** 280–284. DOI: 10.1038/nature24049.

47. Cox, D. B. T. *et al.* (2017) RNA editing with CRISPR-Cas13. *Science* **1027,** 1019–1027. DOI: 10.1126/science.aaq0180.

48. Abudayyeh, O. O. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353,** aaf5573. DOI: 10.1126/science.aaf5573.

49. Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. (2016) Adaptation in CRISPR-Cas Systems. *Mol. Cell* **61,** 797–808. DOI: 10.1016/j.molcel.2016.01.030.

50. Jo, Y. Il, Suresh, B., Kim, H. & Ramakrishna, S. (2015) CRISPR/Cas9 system as an innovative genetic engineering tool: Enhancements in sequence specificity and delivery methods. *Biochim. Biophys. Acta - Rev. Cancer* **1856,** 234–243. DOI: 10.1016/j.bbcan.2015.09.003.

51. Heler, R. *et al.* (2015) Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519,** 199–202. DOI: 10.1038/nature14245.

52. Gao, L. *et al.* (2017) Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35,** 789–792. DOI: 10.1038/nbt.3900.

53. Barrangou, R. & Marraffini, L. A. (2014) CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54,** 234–244. DOI: 10.1016/j.molcel.2014.03.011.

54. Rath, D., Amlinger, L., Rath, A. & Lundgren, M. (2015) The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* **117,** 119–128. DOI: 10.1016/j.biochi.2015.03.025.

55. Cress, B. F. *et al.* (2016) Rapid generation of CRISPR/dCas9-regulated, orthogonally repressible hybrid T7-lac promoters for modular, tuneable control of metabolic pathway fluxes in *Escherichia coli*. *Nucleic Acids Res.* **44,** 4472–4485. DOI: 10.1093/nar/gkw231.

56. Bordoy, A. E., Varanasi, U. S., Courtney, C. M. & Chatterjee, A. (2016) Transcriptional Interference in Convergent Promoters as a Means for Tunable Gene Expression. *ACS Synth. Biol.* **5,** 1331–1341. DOI: 10.1021/acssynbio.5b00223.

57. Gilbert, L. A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451. DOI: 10.1016/j.cell.2013.06.044.

58. Marraffini, L. A. & Sontheimer, E. J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322,** 1843–1845. DOI: 10.1126/science.1165771.

59. Maeder, M. L. *et al.* (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979. DOI: 10.1038/nmeth.2598.CRISPR.

60. Gilbert, L. A. *et al.* (2014) Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 1–15. DOI: 10.1016/j.cell.2014.09.029.

61. Gao, Y. *et al.* (2016) Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. **13,** DOI: 10.1038/nmeth.4042.

62. Zheng, Y. *et al.* (2018) CRISPR interference-based specific and efficient gene inactivation in the brain. *Nat. Neurosci.* **21,** 1. DOI: 10.1038/s41593-018-0077-5.

63. Yao, L., Cengic, I., Anfelt, J. & Hudson, E. P. (2016) Multiple Gene Repression in Cyanobacteria Using CRISPRi. *ACS Synth. Biol.* **5,** 207–212. DOI: 10.1021/acssynbio.5b00264.

64. Cress, B. F. *et al.* (2015) CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in *E. coli*. *ACS Synth. Biol.* **4,** 987–1000. DOI: 10.1021/acssynbio.5b00012.

65. Luo, M. L., Mullis, a. S., Leenay, R. T. & Beisel, C. L. (2014) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43,** 674–681. DOI: 10.1093/nar/gku971.

66. Zuberi, A., Misba, L. & Khan, A. U. (2017) CRISPR Interference (CRISPRi) Inhibition of luxS Gene Expression in E. coli: An Approach to Inhibit Biofil. *Front. Cell. Infect. Microbiol.* **7,** 1–7. DOI: 10.3389/fcimb.2017.00214.

67. Deaner, M. & Alper, H. S. (2017) Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in Saccharomyces cerevisiae. *Metab. Eng.* **40,** 14–22. DOI: 10.1016/j.ymben.2017.01.012.

68. Wang, M., Liu, L., Fan, L. & Tan, T. (2017) CRISPRi based system for enhancing 1-butanol production in engineered Klebsiella pneumoniae. *Process Biochem.* **56,** 139–146. DOI: 10.1016/j.procbio.2017.02.013.

69. Wu, M. Y., Sung, L. Y., Li, H., Huang, C. H. & Hu, Y. C. (2017) Combining CRISPR and CRISPRi Systems for Metabolic Engineering of *E. coli* 1,4-BDO Biosynthesis. *ACS Synth. Biol.* **6,** 2350–2361. DOI: 10.1021/acssynbio.7b00251.

70. Shmakov, S. *et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* **60,** 385–397. DOI: 10.1016/j.molcel.2015.10.008.

71. East-Seletsky, A. *et al.* (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538,** 270–273. DOI: 10.1038/nature19802.

72. East-seletsky, A., Connell, M. R. O., Burstein, D., Knott, G. J. & Jennifer, A. (2017) RNA targeting by functionally orthogonal Type VI-A CRISPR-Cas enzymes. *Mol. Cell* **66,** 373–383.e3. DOI: 10.1016/j.molcel.2017.04.008.

73. Gootenberg, J. S. *et al.* (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356,** 438–442. DOI: 10.1126/science.aam9321.

74. Aman, R. *et al.* (2018) RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biol.* **19,** 1–9. DOI: 10.1186/s13059-017-1381-1.

75. Gu, Z., Biswas, A., Zhao, M. & Tang, Y. (2011) Tailoring nanocarriers for intracellular protein delivery. *Chem. Soc. Rev.* **40,** 3638. DOI: 10.1039/c0cs00227e.

76. Errington, J., Bath, J. & Wu, L. J. (2001) DNA transport in bacteria. *Nat. Rev. Mol. Cell Biol.* **2,** 538–545. DOI: 10.1038/35080005.

77. Ji, W. *et al.* (2014) Specific Gene repression by CRISPRi system transferred through bacterial conjugation. *ACS Synth. Biol.* **3,** 929–931. DOI: 10.1021/sb500036q.

78. Ryan, E. M., Gorman, S. P., Donnelly, R. F. & Gilmore, B. F. (2011) Recent advances in bacteriophage therapy: how delivery routes, formulation, concentration and timing influence the success of phage therapy. *J. Pharm. Pharmacol.* **63,** 1253–64. DOI: 10.1111/j.2042-7158.2011.01324.x.

79. Brasino, M. (2016) Engineering of Filamentous Bacteriophage for Protein Sensing.

80. Bikard, D. *et al.* (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nat. Biotechnol.* 1–6. DOI: 10.1038/nbt.3043.

81. Yosef, I., Manor, M., Kiro, R. & Qimron, U. (2015) Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci.* **112,** 7267–7272. DOI: 10.1073/pnas.1500107112.

82. Milletti, F. (2012) Cell-penetrating peptides: Classes, origin, and current landscape. *Drug Discov. Today* **17,** 850–860. DOI: 10.1016/j.drudis.2012.03.002.

83. Ramakrishna, S. *et al.* (2014) Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res.* **24,** 1020–7. DOI: 10.1101/gr.171264.113.

84. Liu, J., Gaj, T., Patterson, J. T., Sirk, S. J. & Barbas, C. F. (2014) Cell-penetrating peptide-mediated delivery of TALEN proteins via bioconjugation for genome engineering. *PLoS One* **9,** e85755. DOI: 10.1371/journal.pone.0085755.

85. Sun, W. *et al.* (2015) Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing. *Angew. Chemie - Int. Ed.* **54,** 12029–12033. DOI: 10.1002/anie.201506030.

86. Nielsen, P., Egholm, M., Berg, R. & Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254,** 1497–1500. DOI: 10.1126/science.1962210.

87. Nielsen, P. E. (1999) Peptide nucleic acid. A molecule with two identities. *Acc. Chem. Res.* **32,** 624–630. DOI: 10.1021/ar980010t.

88. Egholm, M. *et al.* (1993) PNA hybrdizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365,** 566–568.

89. Nelson, K. E., Levy, M. & Miller, S. L. (2000) Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 3868–71. DOI: 10.1073/PNAS.97.8.3868.

90. Wittung, P., Nielsen, P. E., Buchardt, O., Egholm, M. & Norde´n, B. (1994) DNA-like double helix formed by peptide nucleic acid. *Nature* **368,** 561–563. DOI: 10.1038/368561a0.

91. Courtney, C. & Chatterjee, A. (2014) cis-Antisense RNA and Transcriptional

Interference: Coupled Layers of Gene Regulation. *J. Gene Ther.* **2,** 1–9.

92. Hatamoto, M., Ohashi, A. & Imachi, H. (2010) Peptide nucleic acids (PNAs) antisense effect to bacterial growth and their application potentiality in biotechnology. *Appl. Microbiol. Biotechnol.* **86,** 397–402. DOI: 10.1007/s00253-009-2387-8.

93. Bai, H. & Luo, X. (2010) Antisense Antibacterials: From Proof-Of-Concept to Therapeutic Perspectives. *IntechOpen* 320–344. DOI: 10.5772/33347.

94. Good, L. & Nielsen, P. E. (1998) Antisense inhibition of gene expression in bacteria by PNA targeted to mRNA. *Nat. Biotechnol.* **16,** 355–358. DOI: 10.1038/nbt0498-355.

95. Dryselius, R., Aswasti, S. K., Rajarao, G. K., Nielsen, P. E. & Good, L. (2003) The translation start codon region is sensitive to antisense PNA inhibition in *Escherichia coli*. *Oligonucleotides* **13,** 427–33. DOI: 10.1089/154545703322860753.

96. Nikravesh, A. *et al.* (2007) Antisense PNA accumulates in *Escherichia coli* and mediates a long post-antibiotic effect. *Mol. Ther.* **15,** 1537–1542. DOI: 10.1038/sj.mt.6300209.

97. Courtney, C. M. & Chatterjee, A. (2015) Sequence-specific peptide nucleic acid-based antisense inhibitors of TEM-1 β-Lactamase and mechanism of adaptive resistance. *ACS Infect. Dis.* **1,** 253–263. DOI: 10.1021/acsinfecdis.5b00042.

98. Almeida, C., Azevedo, N. F., Santos, S., Keevil, C. W. & Vieira, M. J. (2011) Discriminating multi-species populations in biofilms with peptide nucleic acid fluorescence in situ hybridization (PNA FISH). *PLoS One* **6,** DOI: 10.1371/journal.pone.0014786.

99. Holtzman, C., Whitney, D., Barlam, T. & Miller, N. S. (2011) Assessment of impact of peptide nucleic acid fluorescence in situ hybridization for rapid identification of coagulase-negative staphylococci in the absence of antimicrobial stewardship interventionδ. *J. Clin. Microbiol.* **49,** 1581–1582. DOI: 10.1128/JCM.02461-10.

100. Englund, E. A. *et al.* (2012) Programmable multivalent display of receptor ligands using peptide nucleic acid nanoscaffolds. *Nat. Commun.* **3,** 614–617. DOI: 10.1038/ncomms1629.

101. Turner, J. J. *et al.* (2007) RNA targeting with peptide conjugates of oligonucleotides, siRNA and PNA. *Blood Cells, Mol. Dis.* **38,** 1–7. DOI: 10.1016/j.bcmd.2006.10.003.

102. Good, L., Awasthi, S. K., Dryselius, R., Larsson, O. & Nielsen, P. E. (2001) Bactericidal antisense effects of peptide-PNA conjugates. *Nat. Biotechnol.* **19,** 360–364. DOI: 10.1038/86753.

103. Kaihatsu, K., Huffman, K. E. & Corey, D. R. (2004) Intracellular uptake and inhibition of gene expression by PNAs and PNA-peptide conjugates. *Biochemistry* **43,** 14340–14347. DOI: 10.1021/bi048519l.

104. Brandén, L. J., Mohamed, A. J. & Smith, C. I. E. (1999) A peptide nucleic acid – nuclear localization signal fusion that mediates nuclear transport of DNA. *Nat. Biotechnol.* **17,**

105. Ma, S. *et al.* (2014) Electroporation-based delivery of cell-penetrating peptide conjugates of peptide nucleic acids for antisense inhibition of intracellular bacteria. *Integr. Biol.* **6,** 973–978. DOI: 10.1039/C4IB00172A.

106. Chiarantini, L. *et al.* (2005) Comparison of novel delivery systems for antisense peptide nucleic acids. *J. Control. Release* **109,** 24–36. DOI: 10.1016/j.jconrel.2005.09.013.

107. Courtney, C. M. (2017) Engineering Synthetic Antibiotics in Non-Traditional Pathways to Counter Antibiotic Resistance.

108. Mellbye, B. L., Puckett, S. E., Tilley, L. D., Iversen, P. L. & Geller, B. L. (2009) Variations in amino acid composition of antisense peptide- phosphorodiamidate morpholino oligomer affect potency against Escherichia coli in vitro and in vivo. *Antimicrob. Agents Chemother.* **53,** 525–530. DOI: 10.1128/AAC.00917-08.

109. Erickson, K. E. (2017) Novel approaches to bioengineering target identifcation: a focus on non-genetic contributions to complex bacterial phenotypes.

110. Veening, J.-W., Smits, W. K. & Kuipers, O. P. (2008) Bistability, Epigenetics, and Bet-Hedging in Bacteria. *Annu. Rev. Microbiol.* **62,** 193–210. DOI: 10.1146/annurev.micro.62.081307.163002.

111. Beaumont, H. J. E., Gallie, J., Kost, C., Ferguson, G. C. & Rainey, P. B. (2009) Experimental evolution of bet hedging. *Nature* **462,** 90–3. DOI: 10.1038/nature08504.

112. Ratcliff, W. C. & Denison, R. F. (2010) Individual-level bet hedging in the bacterium Sinorhizobium meliloti. *Curr. Biol.* **20,** 1740–1744. DOI: 10.1016/j.cub.2010.08.036.

113. Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J. & Hartl, D. L. (2007) Genetic Properties Influencing the Evolvability of Gene Expression. *Science* **317,** 118–122. DOI: 10.1126/science.1140247.

114. Garcia-Bernardo, J. & Dunlop, M. J. (2013) Tunable Stochastic Pulsing in the *Escherichia coli* Multiple Antibiotic Resistance Network from Interlinked Positive and Negative Feedback Loops. *PLoS Comput. Biol.* **9,** e1003229. DOI: 10.1371/journal.pcbi.1003229.

115. Cohen, N. R., Lobritz, M. a & Collins, J. J. (2013) Microbial persistence and the road to drug resistance. *Cell Host Microbe* **13,** 632–42. DOI: 10.1016/j.chom.2013.05.009.

116. Fernández, L., Breidenstein, E. B. M. & Hancock, R. E. W. (2011) Creeping baselines and adaptive resistance to antibiotics. *Drug Resist. Updat.* **14,** 1–21. DOI: 10.1016/j.drup.2011.01.001.

117. Braoudaki, M. & Hilton, A. C. (2004) Adaptive Resistance to Biocides in *Salmonella enterica* and *Escherichia coli* O157 and Cross-Resistance to Antimicrobial Agents. *J. Clin. Microbiol.* **42,** 73–78. DOI: 10.1128/JCM.42.1.73.

118. Fernández, L. & Hancock, R. E. W. (2012) Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clin. Microbiol. Rev.* **25,** 661–81. DOI: 10.1128/CMR.00043-12.

119. Daikos, G. L., Lolans, V. T. & Jackson, G. G. (1991) First-exposure adaptive resistance to aminoglycoside antibiotics in vivo with meaning for optimal clinical use. *Antimicrob. Agents Chemother.* **35,** 117–123. DOI: 10.1128/AAC.35.1.117.

120. Rosner, J. L. (1985) Nonheritable resistance to chloramphenicol and other antibiotics induced by salicylates and other chemotactic repellents in *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 8771–8774. DOI: 10.1073/pnas.82.24.8771.

121. Viveiros, M. *et al.* (2005) Inducement and Reversal of Tetracycline Resistance in *Escherichia coli* K-12 and Expression of Proton Gradient-Dependent Multidrug Efflux Pump Genes. *Antimicrob. Agents Chemother.* **49,** 3578–3582. DOI: 10.1128/AAC.49.8.3578.

122. Reyes, L. H., Almario, M. P., Winkler, J., Orozco, M. M. & Kao, K. C. (2012) Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in *Escherichia coli*. *Metab. Eng.* **14,** 579–590. DOI: 10.1016/j.ymben.2012.05.002.

123. Oz, T. *et al.* (2014) Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. *Mol. Biol. Evol.* **31,** 2387–2401. DOI: 10.1093/molbev/msu191.

124. Ostrowski, E. A., Woods, R. J. & Lenski, R. E. (2008) The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. *Proc. R. Soc. B* **275,** 277–284. DOI: 10.1098/rspb.2007.1244.

125. Toprak, E. *et al.* (2012) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44,** 101–105. DOI: 10.1038/ng.1034.

126. Sánchez-Romero, M. A. & Casadesús, J. (2014) Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 355–60. DOI: 10.1073/pnas.1316084111.

127. Adam, M., Murali, B., Glenn, N. O. & Potter, S. S. (2008) Epigenetic inheritance based evolution of antibiotic resistance in bacteria. *BMC Evol. Biol.* **8,** 52. DOI: 10.1186/1471-2148-8-52.

128. Choi, J. K. & Kim, Y.-J. (2008) Epigenetic regulation and the variability of gene expression. *Nat. Genet.* **40,** 141–7. DOI: 10.1038/ng.2007.58.

129. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. (2003) Noise in eukaryotic gene expression. *Nature* **422,** 633–637. DOI: 10.1038/nature01546.

130. López-Maury, L., Marguerat, S. & Bähler, J. (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* **9,** 583–593. DOI: 10.1038/nrg2398.

131. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. (2002) Stochastic Gene Expression in a Single Cell. *Science* **297,** 1183–1186. DOI: 10.1126/science.1070919.

132. Veening, J.-W. *et al.* (2008) Bet-hedging and epigenetic inheritance in bacterial cell

development. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 4393–8. DOI: 10.1073/pnas.0700463105.

133. White, D. G., Goldman, J. D., Demple, B. & Levy, S. B. (1997) Role of the acrAB locus in organic solvent tolerance mediated by expression of marA, soxS, or robA in *Escherichia coli*. *J. Bacteriol.* **179,** 6122.

134. Rutherford, B. J. *et al.* (2010) Functional genomic study of exogenous n-butanol stress in *Escherichia coli*. *Appl. Environ. Microbiol.* **76,** 1935–1945. DOI: 10.1128/AEM.02323-09.

135. Tamae, C. *et al.* (2008) Determination of antibiotic hypersensitivity among 4,000 single-gene-knockout mutants of *Escherichia coli*. *J. Bacteriol.* **190,** 5981–5988. DOI: 10.1128/JB.01982-07.

136. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. (2007) A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* **130,** 797–810. DOI: 10.1016/j.cell.2007.06.049.

137. Gutierrez, a *et al.* (2013) β-lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. *Nat. Commun.* **4,** 1610. DOI: 10.1038/ncomms2607.

138. Baba, T. *et al.* (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008. DOI: 10.1038/msb4100050.

139. Huang, D. W. *et al.* (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8,** R183. DOI: 10.1186/gb-2007-8-9-r183.

140. Benjamini, Y. & Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **57,** 289–300.

141. Harel, Y. M., Bailone, A. & Bibi, E. (1999) Resistance to bacitracin as modulated by an *Escherichia coli*homologue of the bacitracin ABC transporter BcrC subunit from Bacillus licheniformis. *J. Bacteriol.* **181,** 6176–6178.

142. Wada, T., Hatamoto, Y. & Kutsukake, K. (2012) Functional and expressional analyses of the anti-FlhD4C2 factor gene ydiV in *Escherichia coli*. *Microbiol. (United Kingdom)* **158,** 1533–1542. DOI: 10.1099/mic.0.056036-0.

143. Maciag, A. *et al.* (2011) In vitro transcription profiling of the σs subunit of bacterial RNA polymerase: re-definition of the σs regulon and identification of σs-specific promoter sequence elements. *Nucleic Acids Res.* **39,** 5338–5355. DOI: 10.1093/nar/gkr129.

144. Partridge, J. D. *et al.* (2008) Characterization of the *Escherichia coli* K-12 ydhYVWXUT operon: Regulation by FNR, NarL and NarP. *Microbiology* **154,** 608–618. DOI: 10.1099/mic.0.2007/012146-0.

145. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **7,** 203–214.

146. Mann, C. M. & Markham, J. L. (1998) A new method for determining the minimum inhibitory concentration of essential oils. *J. Appl. Microbiol.* **84,** 538–544. DOI: 10.1046/j.1365-2672.1998.00379.x.

147. Smits, W. K., Kuipers, O. P. & Veening, J.-W. (2006) Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.* **4,** 259–71. DOI: 10.1038/nrmicro1381.

148. Motta, S. S., Cluzel, P. & Aldana, M. (2015) Adaptive resistance in bacteria requires epigenetic inheritance, genetic noise, and cost of efflux pumps. *PLoS One* **10,** 3–8. DOI: 10.1371/journal.pone.0118464.

149. Clinical and Laboratory Standards Institute (2014) *M100-S24. Performance standards for antimicrobial susceptibility testing; Twenty-fourth informational supplement*.

150. De Jong, D. W. & Woodlief, W. G. (1977) Fluorimetric assay of tobacco leaf dehydrogenases with resazurin. *Biochim. Biophys. Acta (BBA)-Enzymology* **484,** 249–259. DOI: 10.1016/0005-2744(77)90081-X.

151. Ho, J. W. K., Stefani, M., dos Remedios, C. G. & Charleston, M. A. (2008) Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics* **24,** i390–i398. DOI: 10.1093/bioinformatics/btn142.

152. Morgan, A. A., Dudley, J. T., Deshpande, T. & Butte, A. J. (2010) Dynamism in gene expression across multiple studies. *Physiol. Genomics* **40,** 128–140. DOI: 10.1152/physiolgenomics.90403.2008.

153. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. a & Gill, R. T. (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* **28,** 856–62. DOI: 10.1038/nbt.1653.

154. Fong, S. S., Joyce, A. R. & Palsson, B. Ø. (2005) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* **15,** 1365–72. DOI: 10.1101/gr.3832305.

155. Sagar, D. M. *et al.* (2017) High-Throughput Block Optical DNA Sequence Identification. *Small* **1703165,** 1703165. DOI: 10.1002/smll.201703165.

156. Abel, G., Korshoj, L. E., Otoupal, P. B., Chatterjee, A. & Nagpal, P. Sequence Identification and Structural Label Mapping of Single RNA Molecules with Quantum Tunneling Spectroscopy. *In Preperation*

157. Zhang, Z., Qian, W. & Zhang, J. (2009) Positive selection for elevated gene expression noise in yeast. *Mol. Syst. Biol.* **5,** 299. DOI: 10.1038/msb.2009.58.

158. Pritchard, C. C., Hsu, L., Delrow, J. & Nelson, P. S. (2001) Project normal: Defining normal variance in mouse gene expression. *Proc. Natl. Acad. Sci.* **98,** 13266–13271.

159. Bar-Even, A. *et al.* (2006) Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38,** 636–643. DOI: 10.1038/ng1807.

160. Dong, T., Yu, R. & Schellhorn, H. (2011) Antagonistic regulation of motility and

transcriptome expression by RpoN and RpoS in *Escherichia coli*. *Mol. Microbiol.* **79,** 375–386. DOI: 10.1111/j.1365-2958.2010.07449.x.

161. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. (2014) Growth rates made easy. *Mol. Biol. Evol.* **31,** 232–238. DOI: 10.1093/molbev/mst187.

162. Ye, J. *et al.* (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13,** 134.

163. Zhou, K. *et al.* (2011) Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. *BMC Mol. Biol.* **12,** 18. DOI: 10.1186/1471-2199-12-18.

164. Tyo, K. E. J., Fischer, C. R., Simeon, F. & Stephanopoulos, G. (2010) Analysis of polyhydroxybutyrate flux limitations by systematic genetic and metabolic perturbations. *Metab. Eng.* **12,** 187–195. DOI: 10.1016/j.ymben.2009.10.005.

165. Takle, G. W., Toth, I. K. & Brurberg, M. B. (2007) Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen Pectobacterium atrosepticum. *BMC Plant Biol.* **7,** 50. DOI: 10.1186/1471-2229-7-50.

166. Bar-Joseph, Z., Gifford, D. K. & Jaakkola, T. S. (2001) Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* **17,** S22-9. DOI: 10.1093/bioinformatics/17.suppl_1.S22.

167. Livak, K. J. & Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−ΔΔCT) Method. *Methods* **25,** 402–8. DOI: 10.1006/meth.2001.1262.

168. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8,** R19. DOI: 10.1186/gb-2007-8-2-r19.

169. Trapnell, C. *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–78. DOI: 10.1038/nprot.2012.016.

170. Anders, S. & Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106. DOI: 10.1186/gb-2010-11-10-r106.

171. Langmead, B. & Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359. DOI: 10.1038/nmeth.1923.

172. Li, H. *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079. DOI: 10.1093/bioinformatics/btp352.

173. Mckenna, A. *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303. DOI: 10.1101/gr.107524.110.

174. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–192. DOI: 10.1093/bib/bbs017.

175. Luria, S. E. & Delbruck, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28,** 491–511.

176. Sarkar, S., Ma, W. T. & Sandri, G. H. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85,** 173–179. DOI: 10.1007/BF00120324.

177. Hall, B. M., Ma, C.-X., Liang, P. & Singh, K. K. (2009) Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25,** 1564–1565. DOI: 10.1093/bioinformatics/btp253.

178. Lamprecht, M. R., Sabatini, D. M. & Carpenter, A. E. (2007) CellProfiler: Free, versatile software for automated biological image analysis. *Biotechniques* **42,** 71–75. DOI: 10.2144/000112257.

179. Trinh, C. T. (2012) Elucidating and reprogramming *Escherichia coli* metabolisms for obligate anaerobic n-butanol and isobutanol production. *Appl. Microbiol. Biotechnol.* **95,** 1083–94. DOI: 10.1007/s00253-012-4197-7.

180. Chen, B., Ling, H. & Chang, M. W. (2013) Transporter engineering for improved tolerance against alkane biofuels in Saccharomyces cerevisiae. *Biotechnol. Biofuels* **6,** 21. DOI: 10.1186/1754-6834-6-21.

181. Jin, C., Yao, M., Liu, H., Lee, C. F. F. & Ji, J. (2011) Progress in the production and application of n-butanol as a biofuel. *Renew. Sustain. Energy Rev.* **15,** 4080–4106. DOI: 10.1016/j.rser.2011.06.001.

182. Dunlop, M. J. (2011) Engineering microbes for tolerance to next-generation biofuels. *Biotechnol. Biofuels* **4,** 32. DOI: 10.1186/1754-6834-4-32.

183. Dürre, P. (2007) Biobutanol: An attractive biofuel. *Biotechnol. J.* **2,** 1525–1534. DOI: 10.1002/biot.200700168.

184. Qureshi, N. & Ezeji, T. C. (2008) Butanol, ‘a superior biofuel’ production from agricultural residues (renewable biomass): recent progress in technology. *Biofuels, Bioprod. Biorefining* **2,** 319–330.

185. Sardessai, Y. & Bhosle, S. (2002) Tolerance of bacteria to organic solvents. *Res. Microbiol.* **153,** 263–268.

186. Xue, C. *et al.* (2014) Integrated butanol recovery for an advanced biofuel: Current state and prospects. *Appl. Microbiol. Biotechnol.* **98,** 3463–3474. DOI: 10.1007/s00253-014-5561-6.

187. Knoshaug, E. P. & Zhang, M. (2009) Butanol tolerance in a selection of microorganisms. *Appl. Biochem. Biotechnol.* **153,** 13–20. DOI: 10.1007/s12010-008-8460-4.

188. Tian, X., Chen, L., Wang, J., Qiao, J. & Zhang, W. (2013) Quantitative proteomics reveals dynamic responses of Synechocystis sp. PCC 6803 to next-generation biofuel butanol. *J. Proteomics* **78,** 326–345. DOI: 10.1016/j.jprot.2012.10.002.

189. Liu, J., Chen, L., Wang, J., Qiao, J. & Zhang, W. (2012) Proteomic analysis reveals

resistance mechanism against biofuel hexane in *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels* **5,** 68. DOI: 10.1186/1754-6834-5-68.

190. Tomas, C. A., Beamish, J., Eleftherios, T. & Papoutsakis, E. T. (2004) Transcriptional Analysis of Butanol Stress and Tolerance in Clostridium acetobutylicum. *J. Bacteriol.* **186,** 2006–2018. DOI: 10.1128/JB.186.7.2006.

191. Wang, S., Dong, S., Wang, P., Tao, Y. & Wang, Y. (2017) Genome Editing in Clostridium saccharoperbutylacetonicum N1-4 with the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **83,** 1–16.

192. Li, Q. *et al.* (2016) CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnol. J.* **11,** 961–972. DOI: 10.1002/biot.201600053.

193. Kaczmarzyk, D., Cengic, I., Yao, L. & Hudson, E. P. (2018) Diversion of the long-chain acyl-ACP pool in Synechocystis to fatty alcohols through CRISPRi repression of the essential phosphate acyltransferase PlsX. *Metab. Eng.* **45,** 59–66. DOI: 10.1016/j.ymben.2017.11.014.

194. Anfelt, J., Hallström, B., Nielsen, J., Uhlén, M. & Hudson, E. P. (2013) Using transcriptomics to improve butanol tolerance of Synechocystis sp. Strain PCC 6803. *Appl. Environ. Microbiol.* **79,** 7419–7427. DOI: 10.1128/AEM.02694-13.

195. Zheng, Y. N. *et al.* (2009) Problems with the microbial production of butanol. *J. Ind. Microbiol. Biotechnol.* **36,** 1127–1138. DOI: 10.1007/s10295-009-0609-9.

196. Nielsen, D. R. *et al.* (2009) Engineering alternative butanol production platforms in heterologous bacteria. *Metab. Eng.* **11,** 262–273. DOI: 10.1016/j.ymben.2009.05.003.

197. Atsumi, S. *et al.* (2008) Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **10,** 305–311. DOI: 10.1016/j.ymben.2007.08.003.

198. Reyes, L. H., Abdelaal, A. S. & Kao, K. C. (2013) Genetic determinants for n-Butanol tolerance in evolved *Escherichia coli* mutants: Cross adaptation and antagonistic pleiotropy between n-butanol and other stressors. *Appl. Environ. Microbiol.* **79,** 5313–5320. DOI: 10.1128/AEM.01703-13.

199. Liu, S. & Qureshi, N. (2009) How microbes tolerate ethanol and butanol. *N. Biotechnol.* **26,** 117–121. DOI: 10.1016/j.nbt.2009.06.984.

200. Fletcher, E., Pilizota, T., Davies, P. R., McVey, A. & French, C. E. (2016) Characterization of the effects of n-butanol on the cell envelope of E. coli. *Appl. Microbiol. Biotechnol.* **100,** 9653–9659. DOI: 10.1007/s00253-016-7771-6.

201. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314,** 1565–8. DOI: 10.1126/science.1131969.

202. Erickson, K. E., Gill, R. T. & Chatterjee, A. (2014) Constrictor: Constraint modification provides insight into design of biochemical networks. *PLoS One* **9,** 1–24. DOI: 10.1371/journal.pone.0113820.

203. Wang, C., Pfleger, B. F. & Kim, S. W. (2017) Reassessing Escherichia coli as a cell factory for biofuel production. *Curr. Opin. Biotechnol.* **45,** 92–103. DOI: 10.1016/j.copbio.2017.02.010.

204. Jones, J. A., Toparlak, T. D. & Koffas, M. A. G. (2015) Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Curr. Opin. Biotechnol.* **33,** 52–59. DOI: 10.1016/j.copbio.2014.11.013.

205. Hsu, P. D., Lander, E. S. & Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157,** 1262–1278. DOI: 10.1016/j.cell.2014.05.010.

206. Li, Y. F. *et al.* (2015) Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metab. Eng.* **31,** 13–21. DOI: 10.1016/j.ymben.2015.06.006.

207. Alonso-Gutierrez, J. *et al.* (2017) Towards industrial production of isoprenoids in *Escherichia coli* : lessons learned from CRISPR-Cas9 based optimization of a chromosomally integrated mevalonate pathway. *Biotechnol. Bioeng.* 1000–1013. DOI: 10.1002/bit.26530.

208. Wu, J., Zhang, X., Xia, X. & Dong, M. (2017) A systematic optimization of medium chain fatty acid biosynthesis via the reverse beta-oxidation cycle in *Escherichia coli*. *Metab. Eng.* **41,** 115–124. DOI: 10.1016/j.ymben.2017.03.012.

209. Cress, B. F., Trantas, E. A., Ververidis, F., Linhardt, R. J. & Koffas, M. A. G. (2015) Sensitive cells: Enabling tools for static and dynamic control of microbial metabolic pathways. *Curr. Opin. Biotechnol.* **36,** 205–214. DOI: 10.1016/j.copbio.2015.09.007.

210. Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. (2003) Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci.* **100,** 9727–32. DOI: 10.1073/pnas.1333928100.

211. Alexander, D. L. *et al.* (2014) Random mutagenesis by error-prone Pol I plasmid replication in *Escherichia coli*. *Methods Mol. Biol.* 31–44. DOI: 10.1385/159259395X.

212. Isken, S. & de Bont, J. A. M. (1998) Bacteria tolerant to organic solvents. *Extremophiles* **2,** 229–238.

213. Pérez, A. *et al.* (2007) Cloning, nucleotide sequencing, and analysis of the AcrAB-TolC efflux pump of Enterobacter cloacae and determination of its involvement in antibiotic resistance in a clinical isolate. *Antimicrob. Agents Chemother.* **51,** 3247–53. DOI: 10.1128/AAC.00072-07.

214. Larson, M. H. *et al.* (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8,** 2180–2196. DOI: 10.1038/nprot.2013.132.

215. Stevenson, G., Andrianopoulos, K., Hobbs, M. & Reeves, P. R. (1996) Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J. Bacteriol.* **178,** 4885–4893.

216. Dunlop, M. J. *et al.* (2011) Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* **7,** 1–7. DOI: 10.1038/msb.2011.21.

217. Park, S. & Lehner, B. (2013) Epigenetic epistatic interactions constrain the evolution of gene expression. *Mol. Syst. Biol.* **9,** 645. DOI: 10.1038/msb.2013.2.

218. Chou, H. H., Delaney, N. F., Draghi, J. a. & Marx, C. J. (2014) Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genet.* **10,** DOI: 10.1371/journal.pgen.1004149.

219. Takatsuka, Y., Chen, C. & Nikaido, H. (2010) Mechanism of recognition of compounds of diverse structures by the multidrug efflux pump AcrB of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **107,** 6559–6565. DOI: 10.1073/pnas.1001460107.

220. Zhu, L., Li, Y. & Cai, Z. (2015) Development of a stress-induced mutagenesis module for autonomous adaptive evolution of *Escherichia coli* to improve its stress tolerance. *Biotechnol. Biofuels* **8,** 1–10. DOI: 10.1186/s13068-015-0276-1.

221. Luhe, A. L., Gerken, H., Tan, L., Wu, J. & Zhao, H. (2012) Alcohol tolerance of *Escherichia coli acrR* and *marR* regulatory mutants. *J. Mol. Catal. B Enzym.* **76,** 89–93. DOI: 10.1016/j.molcatb.2011.11.013.

222. Wiktor, J., Lesterlin, C., Sherratt, D. J. & Dekker, C. (2016) CRISPR-mediated control of the bacterial initiation of replication. *Nucleic Acids Res.* **44,** 1–10. DOI: 10.1093/nar/gkw214.

223. Dunlop, M. J., Keasling, J. D. & Mukhopadhyay, A. (2010) A model for improving microbial biofuel production using a synthetic feedback loop. *Syst. Synth. Biol.* **4,** 95–104. DOI: 10.1007/s11693-010-9052-5.

224. Berens, C., Groher, F. & Suess, B. (2015) RNA aptamers as genetic control devices: The potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol. J.* **10,** 246–257. DOI: 10.1002/biot.201300498.

225. Zalatan, J. G. *et al.* (2014) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160,** 339–350. DOI: 10.1016/j.cell.2014.11.052.

226. Alper, H. & Stephanopoulos, G. (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab. Eng.* **9,** 258–67. DOI: 10.1016/j.ymben.2006.12.002.

227. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. (1991) Long-Term Experimental Evolution in *Escherichia coli*. Adaptation and Divergence During 2,000 Generations. *Am. Nat.* **138,** 1315–1341. DOI: 10.1086/521238.

228. Chatterjee, A. *et al.* (2013) Antagonistic self-sensing and mate-sensing signaling controls antibiotic-resistance transfer. *Proc. Natl. Acad. Sci. USA* **110,** 7086–7090. DOI: 10.1073/pnas.1212256110.

229. Pericone, C. D., Park, S., Imlay, J. A. & Weiser, J. N. (2003) Factors Contributing to Hydrogen Peroxide Resistance in Streptococcus pneumoniae Include Pyruvate Oxidase (SpxB) and Avoidance of the Toxic Effects of the Fenton Reaction Factors Contributing to Hydrogen Peroxide Resistance in Streptococcus pneumoniae I.

DOI: 10.1128/JB.185.23.6815.

230. Raser, J. M. & O'Shea, E. K. (2005) Noise in gene expression: origins, consequences, and control. *Science* **309,** 2010–2013. DOI: 10.1126/science.1105891.

231. Raj, A. & van Oudenaarden, A. (2008) Nature, Nurture, or Chance: stochastic gene expression and its consequences. *Cell* **135,** 216–226. DOI: 10.1016/j.cell.2008.09.050.

232. Kaldalu, N., Mei, R. & Lewis, K. (2004) Killing by Ampicillin and Ofloxacin Induces Overlapping Changes in *Escherichia coli* Transcription Profile. *Antimicrob. Agents Chemother.* **48,** 890–896. DOI: 10.1128/AAC.48.3.890.

233. Ceragioli, M. *et al.* (2010) Comparative transcriptomic and phenotypic analysis of the responses of Bacillus cereus to various disinfectant treatments. *Appl. Environ. Microbiol.* **76,** 3352–60. DOI: 10.1128/AEM.03003-09.

234. Puentes-Téllez, P. E., Kovács, Á. T., Kuipers, O. P. & van Elsas, J. D. (2014) Comparative genomics and transcriptomics analysis of experimentally evolved *Escherichia coli* MC1000 in complex environments. *Environ. Microbiol.* **16,** 856–870. DOI: 10.1111/1462-2920.12239.

235. Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332,** 1193–6. DOI: 10.1126/science.1203801.

236. Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D. & Marx, C. J. (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332,** 1190–2. DOI: 10.1126/science.1203799.

237. Woods, R. J. *et al.* (2011) Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331,** 1433–6. DOI: 10.1126/science.1198914.

238. Sander, J. D. & Joung, J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–55. DOI: 10.1038/nbt.2842.

239. Randall, L. P. & Woodward, M. J. (2002) The multiple antibiotic resistance (mar) locus and its significance. *Res. Vet. Sci.* **72,** 87–93. DOI: 10.1053/rvsc.2001.0537.

240. Duval, V. & Lister, I. M. (2013) MarA , SoxS and Rob of *Escherichia coli* – Global Regulators of Multidrug Resistance , Virulence and Stress Response. *Int. J. Biotechnol. Wellness Ind.* **2,** 101–124.

241. Giese, K. C., Michalowski, C. B. & Little, J. W. (2008) RecA-dependent cleavage of LexA dimers. *J. Mol. Biol.* **377,** 148–61. DOI: 10.1016/j.jmb.2007.12.025.

242. Giraud, A. *et al.* (2001) Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* **291,** 2606–2608. DOI: 10.1126/science.1056421.

243. Galhardo, R. S. *et al.* (2009) DinB upregulation is the sole role of the SOS response in stress-induced mutagenesis in *Escherichia coli*. *Genetics* **182,** 55–68. DOI: 10.1534/genetics.109.100735.

244. Nichols, R. J. *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell* **144,** 143–156. DOI: 10.1016/j.cell.2010.11.052.

245. Soo, V. W. C., Hanson-Manful, P. & Patrick, W. M. (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 1484–1489. DOI: 10.1073/pnas.1012108108.

246. Isalan, M. *et al.* (2008) Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452,** 840–845. DOI: 10.1038/nature06847.

247. Barrick, J. E. *et al.* (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461,** 1243–7. DOI: 10.1038/nature08480.

248. Qi, L. S. *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183. DOI: 10.1016/j.cell.2013.02.022.

249. Phillips, P. C. (2008) Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9,** 855–67. DOI: 10.1038/nrg2452.

250. González, C. *et al.* (2015) Stress-response balance drives the evolution of a network module and its host genome. *Mol. Syst. Biol.* **11,** 827. DOI: 10.15252/msb.20156185.

251. Nevozhay, D., Adams, R. M., van Itallie, E., Bennett, M. R. & Balázsi, G. (2012) Mapping the environmental fitness landscape of a synthetic gene circuit. *PLoS Comput. Biol.* **8,** DOI: 10.1371/journal.pcbi.1002480.

252. Gray, M. J., Wholey, W.-Y. & Jakob, U. (2013) Bacterial responses to reactive chlorine species. *Annu. Rev. Microbiol.* **67,** 141–60. DOI: 10.1146/annurev-micro-102912-142520.

253. Linley, E., Denyer, S. P., McDonnell, G., Simons, C. & Maillard, J.-Y. (2012) Use of hydrogen peroxide as a biocide: new consideration of its mechanisms of biocidal action. *J. Antimicrob. Chemother.* **67,** 1589–96. DOI: 10.1093/jac/dks129.

254. Chopra, I. & Roberts, M. (2001) Tetracycline Antibiotics : Mode of Action , Applications , Molecular Biology , and Epidemiology of Bacterial Resistance Tetracycline Antibiotics : Mode of Action , Applications , Molecular Biology , and Epidemiology of Bacterial Resistance. **65,** DOI: 10.1128/MMBR.65.2.232.

255. Campbell, E. a. *et al.* (2001) Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **104,** 901–912. DOI: 10.1016/S0092-8674(01)00286-0.

256. Andersson, D. I. (2003) Persistence of antibiotic resistant bacteria. *Curr. Opin. Microbiol.* **6,** 452–456. DOI: 10.1016/j.mib.2003.09.001.

257. Fridman, O., Goldberg, A., Ronin, I., Shoresh, N. & Balaban, N. Q. (2014) Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513,** 418–421. DOI: 10.1038/nature13469.

258. Gall, S., Lynch, M. D., Sandoval, N. R. & Gill, R. T. (2008) Parallel mapping of genotypes to phenotypes contributing to overall biological fitness. *Metab. Eng.* **10,** 382–93. DOI: 10.1016/j.ymben.2008.08.003.

259. Gama-Castro, S. *et al.* (2016) RegulonDB version 9.0: High-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* **44,** D133–D143. DOI: 10.1093/nar/gkv1156.

260. Farzadfard, F., Perli, S. D. & Lu, T. K. (2013) Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas.

261. Pomposiello, P. J., Bennik, M. H. J. & Demple, B. (2001) Genome-wide transcriptional profiling of the Escherichia coli responses to superoxide stress and sodium salicylate. *J. Bacteriol.* **183,** 3890–3902. DOI: 10.1128/JB.183.13.3890.

262. Cooper, T. F., Rozen, D. E. & Lenski, R. E. (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **100,** 1072–1077. DOI: 10.1073/pnas.0334340100.

263. Kondrashov, F. a. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. B* **279,** 5048–5057. DOI: 10.1098/rspb.2012.1108.

264. Rolfe, M. D. *et al.* (2012) Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J. Bacteriol.* **194,** 686–701. DOI: 10.1128/JB.06112-11.

265. Rosenberg, E. Y. *et al.* (2003) Bile salts and fatty acids induce the expression of Escherichia coli AcrAB multidrug efflux pump through their interaction with Rob regulatory protein. *Mol. Microbiol.* **48,** 1609–1619. DOI: 10.1046/j.1365-2958.2003.03531.x.

266. Tenaillon, O. *et al.* (2012) The molecular diversity of adaptive convergence. *Science* **335,** 457–61. DOI: 10.1126/science.1212986.

267. Wright, S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Int. Congr. Genet.*

268. Trindade, S. *et al.* (2009) Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet.* **5,** e1000578. DOI: 10.1371/journal.pgen.1000578.

269. World Health Organisation (2013) Global Tuberculosis Report. Available at http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf. Accessed February 5, 2017.

270. Goulart, C. P. *et al.* (2013) Designing Antibiotic Cycling Strategies by Determining and Understanding Local Adaptive Landscapes. *PLoS One* **8,** DOI: 10.1371/journal.pone.0056040.

271. Brown, E. M. & Nathwani, D. (2005) Antibiotic cycling or rotation: A systematic review of the evidence of efficacy. *J. Antimicrob. Chemother.* **55,** 6–9. DOI: 10.1093/jac/dkh482.

272. Romero, P. A. & Arnold, F. H. (2009) Exploring protein fitness landscapes by

directed evolution. *Nat. Rev. Mol. Cell Biol.* **10,** 866–76. DOI: 10.1038/nrm2805.

273. He, X., Qian, W., Wang, Z., Li, Y. & Zhang, J. (2010) Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. *Nat. Genet.* **42,** 272–6. DOI: 10.1038/ng.524.

274. Starr, T. N. & Thornton, J. W. (2016) Epistasis in protein evolution. *Protein Sci.* **25,** 1204–1218. DOI: 10.1002/pro.2897.

275. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. (2012) Epistasis as the primary factor in molecular evolution. *Nature* **490,** 535–538. DOI: 10.1038/nature11510.

276. Palmer, A. C. *et al.* (2015) Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6,** 7385. DOI: 10.1038/ncomms8385.

277. Kester, J. C. & Fortune, S. M. (2014) Persisters and beyond: mechanisms of phenotypic drug resistance and drug tolerance in bacteria. *Crit. Rev. Biochem. Mol. Biol.* **49,** 91–101. DOI: 10.3109/10409238.2013.869543.

278. Kussell, E. & Leibler, S. (2005) Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science (80-.).* **309,** 2075–2078. DOI: 10.1126/science.1114383.

279. Gerdes, S. *et al.* (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J. Bacteriol.* **185,** 5673–5684. DOI: 10.1128/JB.185.19.5673.

280. Szklarczyk, D. *et al.* (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–D452. DOI: 10.1093/nar/gku1003.

281. Gullberg, E. *et al.* (2011) Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathog.* **7,** 1–9. DOI: 10.1371/journal.ppat.1002158.

282. Vila, J. *et al.* (1994) Association between double mutation in gyrA gene of ciprofloxacin- resistant clinical isolates of Escherichia coli and MICs. *Antimicrob. Agents Chemother.* **38,** 2477–2479. DOI: 10.1128/AAC.38.10.2477.

283. Duetz, W. A. *et al.* (2000) Methods for intense aeration, growth, storage, and replication of bacterial strains in microtiter plates. *Appl. Environ. Microbiol.* **66,** 2641–2646. DOI: 10.1128/AEM.66.6.2641-2646.2000.

284. Lobritz, M. A. *et al.* (2015) Antibiotic efficacy is linked to bacterial cellular respiration. *Proc. Natl. Acad. Sci.* **112,** 8173–8180. DOI: 10.1073/pnas.1509743112.

285. Yang, J. H. *et al.* (2017) Antibiotic-Induced Changes to the Host Metabolic Environment Inhibit Drug Efficacy and Alter Immune Function. *Cell Host Microbe* 1–9. DOI: 10.1016/j.chom.2017.10.020.

286. Babu, M. *et al.* (2014) Quantitative Genome-Wide Genetic Interaction Screens Reveal Global Epistatic Relationships of Protein Complexes in Escherichia coli. *PLoS Genet.* **10,** DOI: 10.1371/journal.pgen.1004120.

287. World Health Organisation (2017) *Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. Available at http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/. Accessed April, 2018.*

288. Courtney, C. M. *et al.* (2016) Photoexcited quantum dots for killing multidrug-resistant bacteria. *Nat. Mater.* **15,** 485–588. DOI: 10.1038/nmat4542.

289. Courtney, C. M. *et al.* (2017) Potentiating antibiotics in drug-resistant clinical isolates via stimuli-activated superoxide generation. *Sci. Adv.* **3,** e1701776. DOI: 10.1126/sciadv.1701776.

290. Bognar, A. L., Osborne, C., Shane, B., Singer, S. C. & Ferone, R. (1985) Folylpoly-γ-glutamate synthetase-dihydrofolate synthetase. *J. Biol. Chem.* **260,** 5625–5630.

291. Phillips, G. J. & Silhavy, T. J. (1992) The E. coli ffh gene is necessary for viability and efficient protein export. *Nature* **359,** 744–746. DOI: 10.1038/359744a0.

292. Boross, G. & Papp, B. (2016) No Evidence That Protein Noise-Induced Epigenetic Epistasis Constrains Gene Expression Evolution. *Mol. Biol. Evol.* DOI: 10.1093/molbev/msw236.

293. Li, K. *et al.* (2004) Thioredoxin can influence gene expression by affecting gyrase activity. *Nucleic Acids Res.* **32,** 4563–4575. DOI: 10.1093/nar/gkh794.

294. Cheng, A. A., Ding, H. & Lu, T. K. (2014) Enhanced killing of antibiotic-resistant bacteria enabled by massively parallel combinatorial genetics. *Proc. Natl. Acad. Sci.* **111,** 1400093111-. DOI: 10.1073/pnas.1400093111.

295. Citorik, R. J., Mimee, M. & Lu, T. K. (2014) Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* 1–7. DOI: 10.1038/nbt.3011.

296. Gomaa, A. a *et al.* (2014) Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR/Cas Systems. *MBio* **5,** e00928-13. DOI: 10.1128/mBio.00928-13.Editor.

297. Eriksson, M., Nielsen, P. E. & Good, L. (2002) Cell permeabilization and uptake of antisense peptide-peptide nucleic acid (PNA) into Escherichia coli. *J. Biol. Chem.* **277,** 7144–7147. DOI: 10.1074/jbc.M106624200.

298. Berenbau, M. C., Yu, V. L. & Felegie, T. P. (1983) Synergy with double and triple antibiotic combinations compared. *J. Antimicrob. Chemother.* **12,** 555–563.

299. Kaelin, W. G. (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat. Rev. Cancer* **5,** 689–98. DOI: 10.1038/nrc1691.

300. Davies, S. C. (2013) Infections and the Rise of Antimicrobial Resistance. *Annu. Rep. Chief Med. Off. - Vol. Two* **Chapter 5.,** 73. DOI: 10.1016/S0140-6736(13)60604-2.

301. World Economic Forum (2014) *Global Risks 2014, Ninth Edition*.

302. O'Neill, J. (2014) Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Review on antimicrobial resistance.

303. Smith, P. a & Romesberg, F. E. (2007) Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation. *Nat. Chem. Biol.* **3,** 549–56. DOI: 10.1038/nchembio.2007.27.

304. Abeylath, S. C. & Turos, E. (2008) Drug delivery approaches to overcome bacterial resistance to beta-lactam antibiotics. *Expert Opin. Drug Deliv.* **5,** 931–949. DOI: 10.1517/17425247.5.9.931.

305. Yeh, P., Tschumi, A. I. & Kishony, R. (2006) Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* **38,** 489–494. DOI: 10.1038/ng1755.

306. Torella, J. P., Chait, R. & Kishony, R. (2010) Optimal drug synergy in Antimicrobial Treatments. *PLoS Comput. Biol.* **6,** 1–9. DOI: 10.1371/journal.pcbi.1000796.

307. Cokol, M. *et al.* (2011) Systematic exploration of synergistic drug pairs. *Mol. Syst. Biol.* **7,** 1–9. DOI: 10.1038/msb.2011.71.

308. Zhou, A. *et al.* (2015) Synergistic interactions of vancomycin with different antibiotics against Escherichia coli: Trimethoprim and nitrofurantoin display strong synergies with vancomycin against wild-type E. coli. *Antimicrob. Agents Chemother.* **59,** 276–281. DOI: 10.1128/AAC.03502-14.

309. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. (2010) How antibiotics kill bacteria: From targets to networks. *Nat. Rev. Microbiol.* **8,** 423–435. DOI: 10.1038/nrmicro2333.

310. Hegreness, M., Shoresh, N., Damian, D., Hartl, D. & Kishony, R. (2008) Accelerated evolution of resistance in multidrug environments. *Proc. Natl. Acad. Sci.* **105,** 13977–13981. DOI: 10.1073/pnas.0805965105.

311. Pena-Miller, R. *et al.* (2013) When the Most Potent Combination of Antibiotics Selects for the Greatest Bacterial Load: The Smile-Frown Transition. *PLoS Biol.* **11,** 14–16. DOI: 10.1371/journal.pbio.1001540.

312. Lee, S. *et al.* (2009) Targeting a bacterial stress response to enhance antibiotic action. *Proc. Natl. Acad. Sci.* **106,** 14570–14575. DOI: 10.1073/pnas.0903619106.

313. Liu, A. *et al.* (2010) Antibiotic sensitivity profiles determined with an Escherichia coli gene knockout collection: Generating an antibiotic bar code. *Antimicrob. Agents Chemother.* **54,** 1393–1403. DOI: 10.1128/AAC.00906-09.

314. Weiss, S. J., Mansell, T. J., Mortazavi, P., Knight, R. & Gill, R. T. (2016) Parallel Mapping of Antibiotic Resistance Alleles in Escherichia coli. 1–18. DOI: 10.1371/journal.pone.0146916.

315. Baym, M., Stone, L. K. & Kishony, R. (2016) Multidrug evolutionary strategies to reverse antibiotic resistance. *Science (80-.).* **351,** DOI: 10.1126/science.aad3292.

316. Nichols, R. J. *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell* **144,** 143–156. DOI: 10.1016/j.cell.2010.11.052.

317. Cottarel, G. & Wierzbowski, J. (2007) Combination drugs, an emerging option for antibacterial therapy. *Trends Biotechnol.* **25,** 547–555. DOI:

10.1016/j.tibtech.2007.09.004.

318. Cirz, R. T. *et al.* (2005) Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol.* **3,** 1024–1033. DOI: 10.1371/journal.pbio.0030176.

319. Courtney, C. M. *et al.* (2018) Rationally designed antibiotics against non-traditional genes and pathways eliminate multi-drug resistant bacteria. *In Preperation*

320. O'Leary, N. A. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44,** D733-45. DOI: 10.1093/nar/gkv1189.

321. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25. DOI: 10.1186/gb-2009-10-3-r25.

322. Bikard, D. *et al.* (2014) Development of sequence-specific antimicrobials based on programmable CRISPR-Cas nucleases. *Nat. Biotechnol.* **32,** 1146–1150. DOI: 10.1530/ERC-14-0411.Persistent.

323. White, R. *et al.* (2011) Holin triggering in real time. *Proc. Natl. Acad. Sci.* **108,** 798–803. DOI: 10.1073/pnas.1011921108.

324. Gründling, A., Manson, M. D. & Young, R. (2001) Holins kill without warning. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 9348–9352. DOI: 10.1073/pnas.151247598.

325. Smith, D. L., Struck, D. K., Scholtz, J. M. & Young, R. (1998) Purification and biochemical characterization of the lambda holin. *J. Bacteriol.* **180,** 2531–2540.

326. Savva, C. G. *et al.* (2008) The holin of bacteriophage lambda forms rings with large diameter. *Mol. Microbiol.* **69,** 784–793. DOI: 10.1111/j.1365-2958.2008.06298.x.

327. Grundling, A., Smith, D. L., Blasi, U. & Young, R. (2000) Dimerization between the holin and holin inhibitor of phage λ. *J. Bacteriol.* **182,** 6075–6081. DOI: 10.1128/JB.182.21.6075-6081.2000.

328. Marr, A. G. (1991) Growth rate of *Escherichia coli*. *Microbiol. Rev.* **55,** 316–333.

329. Purcell, O., Grierson, C. S., Bernardo, M. & Savery, N. J. (2012) Temperature dependence of ssrA-tag mediated protein degradation. *J Biol Eng* **6,** 10. DOI: 10.1186/1754-1611-6-10.

330. Hayes, R. P. *et al.* (2016) Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* **530,** 499–503. DOI: 10.1038/nature16995.

331. Dominguez, A. A., Lim, W. A. & Qi, L. S. (2016) Beyond editing: Repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17,** 5–15. DOI: 10.1038/nrm.2015.2.

332. Reynolds, A. *et al.* (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.* **22,** 326–330. DOI: 10.1038/nbt936.

333. Laursen, B. S., Sørensen, H. P., Kusk, K., Sperling-petersen, H. U. & Mortensen, K. K. (2005) Initiation of Protein Synthesis in Bacteria Initiation of Protein Synthesis in Bacteria. **69,** DOI: 10.1128/MMBR.69.1.101.

334. Simonetti, A. *et al.* (2008) Structure of the 30S translation initiation complex. *Nature* **455,** 416–420. DOI: 10.1038/nature07192.

335. O'Connell, M. R. *et al.* (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516,** 263–266. DOI: 10.1038/nature13769.

336. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.* **36,** 70–74. DOI: 10.1093/nar/gkn188.

337. Liang, P. *et al.* (2015) CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell* **6,** 363–372. DOI: 10.1007/s13238-015-0153-5.

338. Kang, X., He, W., Huang, Y., Yu, Q. & Chen, Y. (2016) Introducing precise genetic modifications into human 3PN embryos by CRISPR / Cas-mediated genome editing. *J. Assist. Reprod. Genet.* **33,** 581–588. DOI: 10.1007/s10815-016-0710-8.

339. Callaway, E. (2016) Second Chinese team reports gene editing in human embryos. *Nature* 4–7. DOI: 10.1038/nature.2016.19718.

340. Sugarman, J. (2015) Ethics and germline gene editing. *EMBO Rep.* **16,** 879–880. DOI: 10.15252/embr.201540879.

341. Mulvihill, J. J. *et al.* (2017) Ethical issues of CRISPR technology and gene editing through the lens of solidarity. *Br. Med. Bull.* **122,** 17–29. DOI: 10.1093/bmb/ldx002.

342. Kohn, D. B., Porteus, M. H. & Scharenberg, A. M. (2016) Ethical and regulatory aspects of genome editing. *Blood* **127,** 2553–2560. DOI: 10.1182/blood-2016-01-678136.

343. Lanphier, E., Urnov, F., Haecker, S. E., Werner, M. & Smolenski, J. (2015) Don't edit the human germ line. *Nature* **519,** 410–411. DOI: 10.1038/519410a.

344. Ma, H. *et al.* (2017) Correction of a pathogenic gene mutation in human embryos. *Nature* **548,** 413–419. DOI: 10.1038/nature23305.