Methods to study protein folding and evolution *in vivo*

by

Christopher Lawrence Moore
B.S. in Chemistry (2012)
University of Texas at Dallas, Richardson, TX

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

At the

Massachusetts Institute of Technology

June 2018

Signature redacted

Signature of Author: _____
Department of Chemistry
April 5, 2018

Signature redacted

Certified by: _____
Matthew D. Shoulders
Whitehead Career Development Associate Professor
Thesis supervisor

Signature redacted

Accepted by: _____
Robert W. Field
Chairman, Department Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Signature redacted

Catherine L. Drennan
Professor of Chemistry and Biology
HHMI Professor and Investigator
Department of Chemistry and Department of Biology, MIT
Thesis committee chair

Signature redacted

Matthew D. Shoulders
Whitehead Career Development Associate Professor
Department of Chemistry, MIT
Thesis supervisor

Signature redacted

Bradley L. Pentelute
Pfizer-Laubach Career Development Professor
Department of Chemistry, MIT
Thesis committee member

Methods to study protein folding and evolution *in vivo*

by

Christopher Lawrence Moore

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

Abstract

Proteins are complex biomolecules that fold into distinct shapes to carry out functions which are integral to biological processes. As it turns out, many proteins readily misfold after translation, and failure to address protein misfolding in the context of a living system can have severely detrimental biological consequences.

Living systems have developed complex gene networks consisting of chaperones and quality control factors which maintain protein homeostasis (proteostasis) by actively monitoring protein folding processes in an organelle-specific and dealing with protein misfolding in a stress-responsive manner. While we know that these "proteostasis networks" are capable of influencing protein folding, we lack molecular details regarding how particular components of proteostasis networks work in concert to deal with protein misfolding. Unfortunately, this gap in knowledge also prevents us from understanding the consequences of proteostasis regulation on higher order biological processes, such as the impact chaperone and quality control factors have on protein evolution. Furthermore, we are not able to comment on how dysfunctional proteostasis networks contribute to prominent disease states, including neurodegeneration, cancer, and even pathogenic infections. Thus, an overarching interest in the Shoulders lab at MIT is to fill the aforementioned knowledge gaps by studying how living metazoan systems handle protein folding problems.

The lack of available methods for controlling the activity of proteostasis network components has significantly limited the study of proteostasis in metazoans. In this thesis, I present work that has focused on addressing the limitation in chemical biology tools for studying proteostasis by developing chemical genetic methods tune the level of proteostasis components. Similarly, the inability to conveniently explore protein folding and fitness landscapes on the laboratory timescale has hindered the study of evolution in higher eukaryotes. Thus, my later work sought to overcome this limitation by creating new evolution platforms. Though the inspiration for my work stemmed from a desire to study proteostasis and evolution in metazoans, the methods I developed have allowed other scientists to overcome technical limitations in their own work and progressed the study of many other biological processes beyond proteostasis.

Thesis Supervisor: Matthew D. Shoulders
Title: Whitehead Career Development Associate Professor

Acknowledgements

Graduate school has been a crazy, incredible trip. I'll be the first to admit it has been a journey full of ups and downs and certainly wasn't a walk in the park. With that in mind, one thing is very clear: I would not have made it to this point without the kindness, support and generosity I received from so many people along the way.

First of all, I need to thank my advisor Dr. Matthew Shoulders. It has been an incredible experience to help establish your lab and watch it grow into the small, bustling community that it is today. You've really done an incredible job fostering an environment that not only encourages the development of great scientists, but also great people who support others in their goals. From the first day I met you and talked about science along with what you envisioned, I could sense that learning under your tutelage was going to be a once-in-a-lifetime opportunity. Needless to say, you did not disappoint. You pushed me to work harder, to exceed my limits, to think outside the box and to never be scared to access creative solutions. You were there with measured guidance during the best of times and patient support during the worst of times. In the end, you demanded that I grow into the best scientist, and person, that I could be. For all that you have given me as a mentor and friend, I am forever thankful.

I would also like to thank my thesis committee members, Prof. Cathy Drennan and Brad Pentelute. Cathy, I can say that every exchange I have had with you has been a positive experience, whether it was TAing, updating you on the latest science or just talking about life. I am so grateful for your honest and valuable input. Brad, what's up man! I made it! You have always been a bundle of joy and encouraged me to shoot for the moon. I don't think I can count the number of times you've brightened my day or provided valuable input with just a short exchange of words in passing while in transit. Lastly, thank you Jim for your valuable input in oral examinations and for opening your lab to myself and my labmates. I am so glad you have taken the time to be a member of my committee.

Of course, I have to thank my labmates in the Shoulders Lab who I have had the pleasure of working with. You all are so insanely awesome, and I don't think I can convey how lucky I am to have experienced all your support and guidance. I can honestly say it has been an incredible opportunity to learn from impart wisdom to many of you. To the directed evolution team, thank you for your hard work and for inviting me to be a part of the development process (#ForeverHADES). To Luke, Daniel, Sebastian and Nancy, I wish you the best and know you will be successful if you apply the same work ethic you did for me as mentees. To Rebecca, thank you for the opportunity to closely collaborate and for pushing the proteostasis tools to the next level. Louis, thanks for hanging in there with your incredible patience as we journeyed through the fire and flames of assay optimization and continuous culture experiments in the warm room to develop an exciting new technology. Chris R. and Sam, your fun and stimulating side conversations will always be a treasured distraction. To everybody – Andrew, Emmanuel, Mahender, Patreece, Becca, Pyae, Madeline, Angela, Chet, Duc, Chris R., Louis, Azade, Kenny, Rebecca, Sam, Chichi Michelle, Alex, Sebastian, Nancy, Luke, Agata, Luna, Apolonia, Nora, Diya, Daniel, Chyleigh, Helen, Eileen and Chris H. – thank you all so much!

Also, I have to thank my family and friends for their amazing support. Tom, Jon, and Lauren, thank you for listening me talk about crazy ideas and for dealing with my ups and downs over the years. Shout out to "the Possum" and friends for your incredible support through the ups and the downs of life.

Michelle, Matthew, Cinzia and Johnny, thank you for welcoming me into your lives with warm love and support. I'm excited to officially be a part of the family, but in truth, I have felt a part of it for quite some time now. Thank you for making me feel so welcome and loved, I truly appreciate it.

Mom and Oscar, thank you for your patience and love. I would not be where I am today without all that you have done for me, through all of the ups and downs. Your surprise visits were always a blast and I can't tell you how nice it was to take those breaks, whether it was just a phone call or a night on the town. Tash, you're the best sibling a brother could ask for. I'm so proud of you and all that you've accomplished.

And finally, I have to thank my other half, Chiara. I honestly don't even know where to start. You have been in my corner from day one. You truly are an amazing source of inspiration and I can't believe I have found somebody who is so wonderful to be my best friend. Thank you for all that you give, for all of your patience, and for all of your love.

Dedication

This thesis is dedicated to

A mom that raised me to do anything I put my mind to,

A dad who gave me priceless memories,

A sister who can always make me laugh no matter what,

A stepfather that showed me that patience truly is a virtue,

Friends that have always had my back through thick and thin,

And a wife who brings me endless joy and happiness.

Thank you all so much for your love, kindness, and generosity.

In memory of my dad, Gilbert Gomez

Not a day goes by when you are not loved and missed

Table of Contents

List of Figures

Chapter 1: Progress, challenges, and opportunities in the development of methods to study proteostasis

Chapter 2: A Transportable, Chemical Genetic Methodology for the Small Molecule-Mediated Inhibition of Heat Shock Factor 1

Chapter 3: Multidimensional chemical control of CRISPR-Cas9

Chapter 4: A Processive Protein Chimera Introduces Mutations Across Defined DNA Regions *in vivo*

List of Tables

Abbreviations

4OHT	(*Z*)-4-hydroxytamoxifen
µg/mL	micrograms per a milliliter
Akt	Protein kinase B
Amp	ampicillin
ANOVA	Analysis of variance
ara	arabinose
Ars	arsenite
ASCL1	Achaete-Scute Family BHLH Transcription Factor 1
C	Celsius
Cas9	CRISPR-associated protein 9
cHSF1	constitutively active dominant positive heat shock factor 1
CTH	Cystathionine gamma-lyase
dCas9	deactivated CRISPR-associated protein 9
dSaCas9	deactivated Staphylococcus aureus CRISPR-associated protein 9
dSpCas9	deactivated Streptococcus pyogenes CRISPR-associated protein 9
cDNA	complimentary DNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
d	days
DBD	DNA-binding domain
DD	destabilizing domain
DHFR	dihydrofolate reductase
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
DNAJB	DnaJ homolog subfamily B member protein
n-cHSF1	dominant negative heat shock factor 1 with constitutive activity
dox	doxycycline
DPF2	Double PHD Fingers 2
drApo1	deactivated rApo1
DTT	dithothreitol
E. coli	Escherichia coli
*ec*DHFR	Escherichia coli dihydrofolate reductase
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
EMS	ethyl methanesulfonate
EMX1	Empty Spiracles Homeobox 1
ER50	Destabilizing Domains derived from the Human Estrogen Receptor
ERDJ	ER-Associated DnaJ protein
ERLBD	estrogen receptor ligand binding domain
ERO1	ER oxidoreductin
ERO1L	ERO1-like alpha protein gene
ERO1LB	ERO1-like beta precursor protein gene
FACS	fluorescence-assisted cell sorting
FBS	fetal bovine serum
FDR	false discovery rate
FKBP	FK506-Binding Protein
FLuc	firefly luciferase
Fos	fosfomycin
h	hours

GFP	green fluorescent protein
GO	gene ontology
GRP94	HSP90 homologue residing in endoplasmic reticulum
HEK293	human embryonic kidney cells 293
HEK293T-REx	human embryonic kidney cells 293 with Tet Repressor Expression
$HEK293^{HSR}$	human embryonic kidney cells 293 with heat shock response control
$HEK293^{DD\text{-}HSR}$	HEK293 with heat shock response control through destabilizing domains
HMOX1	Heme Oxygenase 1
HSF1	Heat Shock Factor 1
HSP	heat shock protein
HSPA1A	Heat Shock Protein Family A (HSP70) Member 1A
HSP40	heat shock protein 40
HSP40/70	heat shock protein 40 and heat shock protein 70 subnetwork
HSP70	heat shock protein 40
HSP90	heat shock protein 90
HSP90AA1	heast shock protein 90 family A-Alpha gene
HSP104	heat shock protein 104
HSP110	heat shock protein 110
HSPH1	gene name for HSP110
HSR	heat shock response
HYOU1	hypoxia upregulated protein 1
IL1RN	Interleukin 1 receptor gene
Kan	kanamycin
kDa	kilodalton
Lac	lactose
LacI	Lac repressor
lacO	LacI operon
min	minute
mL	milliliter
MP6	mutagenesis plasmid generation 6
MS2	RNA-binding MS2 phage coat protein
MT1F	metallothionein 1F
MT1X	metallothionein 1X
MutaT7	protein chimera of rApo1 and T7 RNA polymerase
NANOG	homeobox protein
nM	nanomolar
o/n	overnight
p65	NF-kappa-B p65 subunit transcriptional activation domain
PACE	phage-assisted continuous directed evolution
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PEST	degradation tag for protein
PP7	RNA-binding PP7 phage coat protein
Proteostasis	protein homeostasis
qPCR	quantitative polymerase chain reaction
RT-PCR	reverse transcription polymerase chain reaction
RD	regulatory domain
Rif	rifampicin
RIPA	radioimmunoprecipitation assay buffer
RNA	ribonucleic acid

RNP	ribonucleoprotein
RPLP2	Ribosomal Protein Lateral Stalk Subunit P2
rpm	revolutions per minute
rpsL	ribosomal protein subunit L of E. coli
rt	room temperature
rApo1	rat apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1
SaCas9	Staphylococcus aureus CRISPR-associated protein 9
sd	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sem	standard error of the mean
sgRNA	single guide RNA
SLC30A2	Solute Carrier Family 30 Member 2 (zinc transporter)
SpCas9	deactivated Streptococcus pyogenes CRISPR-associated protein 9
Strep	streptomycin
T	temperature
T7	T7 RNA polymerase
T7-pol	T7 RNA polymerase
TAD	transcriptional activation domain
TALE	transcriptional activatior-like effector
TALEN	transcriptional activatior-like effector nuclease
tdTomato	tandem dimeric tomate red fluorescent protein
Tet	tetracycline
TMP	trimethoprim
ung	DNA uracil glycosylase gene
v/v	volume per volume fraction
VEGFA	Vascular endothelial growth factor A
Veh	vehicle
VP16	viral protein 16 transcriptional activation domain
VP64	viral protein 64 transcriptional activation domain
VP192	viral protein 192 transcriptional activation domain
Zeo	zeocin
ZNF	zinc finger

Chapter 1: Progress, challenges, and opportunities in the development of methods to study proteostasis

1.1 Introduction to Protein Homeostasis in the Cytosol and Nucleus

1.1.1 Protein Folding in Living Systems

Proteins are a major class of biomolecules that carry out a diverse array of reactions, scaffolding functions, and many other processes which are required for life. While proteins are initially synthesized from a pool of 20 amino acids into long linear chains, most of these biomolecules become functional by folding into a defined three-dimensional shape. Thus, the process by which proteins achieve a functional form, termed "protein folding", is incredibly important for maintaining living systems.

Unfortunately, the protein folding process is inherently complex and imperfect (**Figure 1.1**)[1-3]. Even in a well-controlled experimental setting, proteins have a propensity to enter an ensemble of conformations instead of exclusively folding into a functional form[4, 5]. Some of these conformations are non-functional and may be states of terminal misfolding if the protein becomes energetically locked in a kinetically trapped conformation. Some conformations may have an increased propensity to interact with and stick to other proteins, leading to oligomerization and ultimately the formation of proteinaceous aggregates.

In living systems, protein misfolding and aggregation is exacerbated by molecular crowding[6]. Some estimates list macromolecular crowding at approximately 30% of the total cell's volume. While this exclusion of volume constrains space and promotes folding, the fact that the crowding agent is composed of many other proteins with hydrophobic stretches also trying to fold is problematic. The high concentration of proteins entering partially folded or misfolded conformations creates a high degree of heterogeneity in terms of hydrophobicity and promotes non-specific intermolecular interactions that promote aggregation. Furthermore, a living system with complex organellular structures may also synthesize a protein in one location yet depend on the protein to carry out an intended function in a separate compartment. In summary, protein folding is an incredibly daunting challenge that must be overcome in cells.

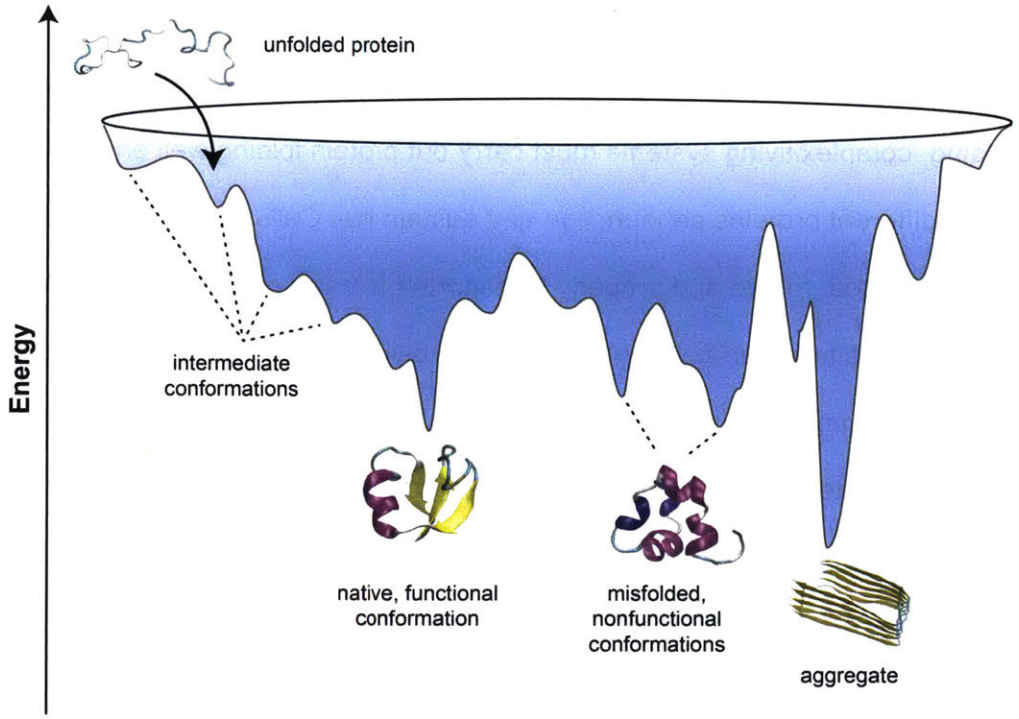

Figure 1.1 | Protein folding is a complex and imperfect process

Theoretical diagram showing a representative energy landscape diagram of the protein folding process. Each conformation a protein enters alters intramolecular interactions, which will subsequently change the free energy of the protein. The relative energy of each conformation is represented as a point, and the numerous conformations result in a continuous surface forming a funnel. Particular conformations represent states of minimal energy and can correspond to functional native conformation or nonfunctional/misfolded conformations. Other low energy state conformations can promote intermolecular interactions and result in aggregation.

1.1.2 Protein Homeostasis

Though the crowded, heterogeneous, and dynamic cellular environment is not ideal for protein folding, complex living systems must carry out protein folding well enough such that thousands of different proteins can function and sustain life. Cells must ensure that enough protein is synthesized, folded and properly transported to carry out required cellular functions. The state of balance between cellular processes involved in maintaining functional protein levels is termed "protein homeostasis", or "proteostasis"[7, 8]. In order to maintain proteostasis, living systems have to contend with protein misfolding, aggregation, and mistrafficking. As each of these deleterious events are exacerbated by physical and chemical stresses that are constantly encountered in the environment, living systems must tightly regulate proteostasis in a highly dynamic and rapidly responsive manner. Cells have evolved to adaptively handle protein folding problems by developing complex gene networks consisting of chaperones, quality control factors, and stress-responsive transcription factors that are dedicated to maintaining proteostasis (**Figure 1.2**).

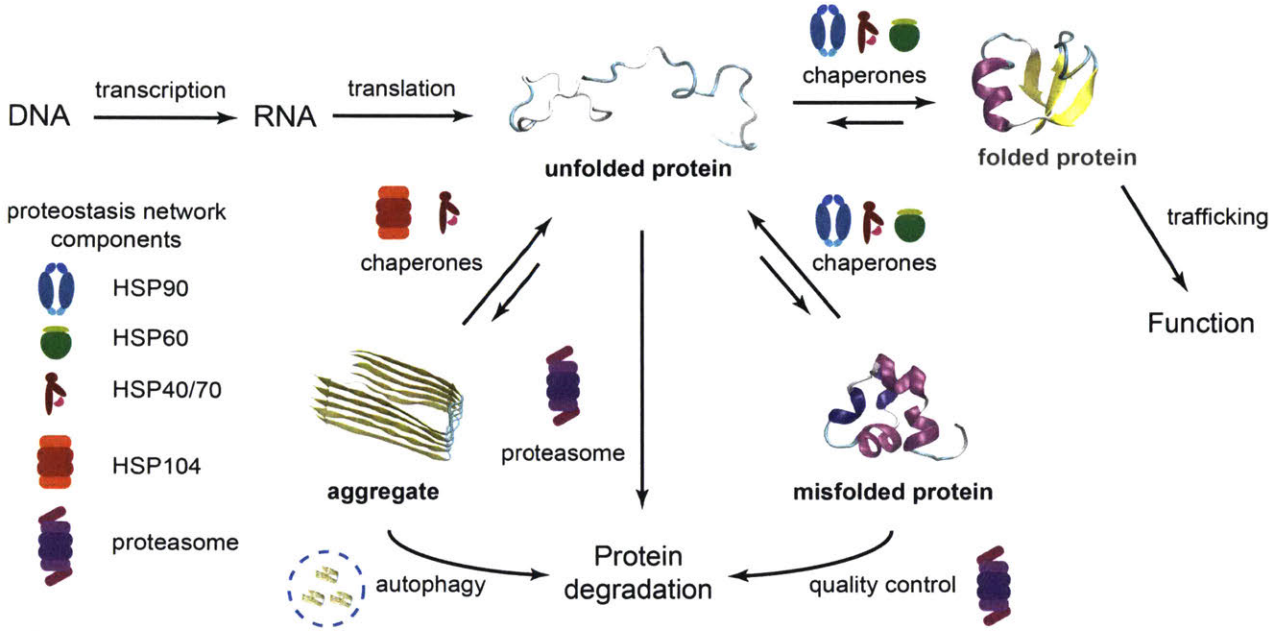

Figure 1.2 | Protein folding in living systems is mediated by proteostasis networks

Diagram demonstrating the different pathways a protein can access following translation. Living systems possess a number of proteostasis network components which regulate and modify flux through these different pathways. Some major components, including heat shock proteins (HSP40, HSP70, HSP90 and HSP104) and degradation components (proteasome), are highlighted next to protein folding processes they are involved in.

1.2 Chemical Biology Approaches to Perturb Cytosolic and Nuclear Proteostasis

1.2.1 Conventional Approaches to Study Proteostasis are Informative, But Technically Limited

A plethora of experimental work until now has identified chaperones, quality control factors and transcription factors which comprise the metazoan cytosolic and nuclear proteostasis networks. Central players that aid in the protein folding processes of these compartments consist of homologues of heat shock proteins (HSP40/70 and HSP90) and other chaperones and co-chaperones that actively bind to client proteins[9, 10]. The ubiquitin-proteasome system performs quality control by recognizing and degrading misfolded proteins[11, 12]. Importantly, Heat Shock Factor 1 (HSF1) is the stress-responsive transcription factor that acts as the master regulator to determine how much of the aforementioned chaperones and quality control components are expressed (**Figure 1.3**)[13, 14].

While the roles that some components play in the cytosolic proteostasis network have been identified, we still lack molecular details regarding how these and many other components work in concert to actively address protein misfolding[15, 16]. This gap in knowledge not only means we have a limited fundamental understanding of protein folding in cells, but also that we do not know how proteostasis networks can be utilized to address neurodegenerative diseases, cancer, and other disorders related to dysregulated proteostasis[7, 17]. Furthermore, we lack an understanding regarding the consequences of proteostasis regulation for higher order biological processes, such as the impact chaperone and quality control factors may have on protein evolution.

A significant reason for the current lack of understanding concerning proteostasis network components in the cytosol (but also in other compartments) lies in the limited methods which are available to create perturbations within proteostasis networks. Initial approaches to study proteostasis networks have relied on genetic overexpression or knockout[18]. While these methods have been utilized with some success up until now, they can be difficult to apply in many systems, particularly if the proteostasis component of interest is an essential gene.

Furthermore, overexpression methods are inherently not tunable, so resulting expression levels will typically not match what is observed naturally. As chaperone activity has been known to vary considerably in response to different stoichiometric balances with co-chaperones and other cellular factors[15, 18], it is important to consider how overexpression results may relate to endogenous expression of proteostasis components. Due to these technical limitations, more recent efforts to perturb proteostasis have focused on chemical biology strategies which enable user-mediated, tunable, and scalable control of proteostasis networks via small molecule treatments.

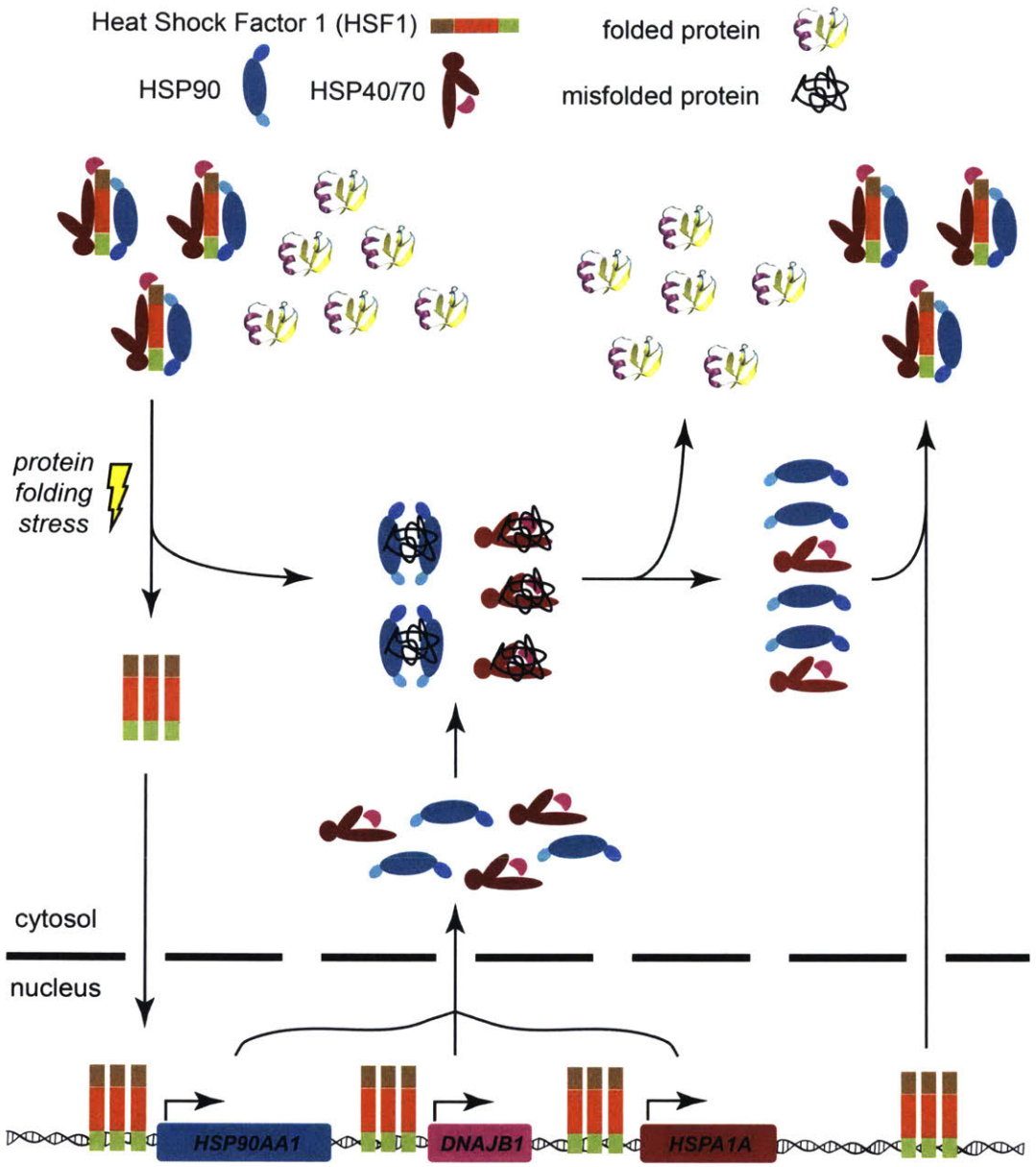

Figure 1.3 | Heat Shock Factor 1 dynamically alters cytosolic proteostasis network

Schematic illustrating the stress-responsive activity of Heat Shock Factor 1 (HSF1). Protein folding stress results in misfolding of chaperone clients. HSF1 is initially bound to chaperones and held in an inactive state until chaperones disengage with HSF1 and engage misfolded clientele. Subsequent trimerization of HSF1, translocation to the nucleus, and binding to DNA results in transcriptional upregulation of proteostasis network components that can directly address protein misfolding. After misfolding is addressed, free chaperones inactivate HSF1.

1.2.2 Chaperone, Proteasome, and Transcription Factor Modulation Through Chemical Biology

More recent methodologies have utilized small molecule modulators of proteostasis network components to cause perturbations of protein folding *in vivo* and study the associated outcomes. One of the most successful small molecule targeting approaches involves inhibitors of chaperones, particularly the HSP90 family[19-26]. HSP90s are a highly conserved class of chaperones that function as a dimeric ATPase complex to stabilize and aid in the folding of client proteins. Approaches to inhibit HSP90 have focused on the identification of parent compounds that typically bind to an ATPase domain or some other regulatory region of the chaperone. Subsequent pharmacological development of these lead compounds has led to extremely potent and selective inhibitors, some of which have even demonstrated specificity to particular closely related paralogs (and even isoforms) of HSP90[26, 27]/ Similarly, small molecule-based inhibitors of HSP70 have enabled pioneering studies of the HSP40/70 inhibitor in living living systems[28-32]. While the development of chaperone inhibitor compounds has been experimentally beneficial in progressing what we know about these particular chaperones, implementation of these inhibitory molecules typically causes significant protein misfolding[33]. For cytosolic HSP90, this protein misfolding causes stress that activates HSF1, which subsequently upregulates all other chaperones in the system (**Figure 1.4**). As most experiments utilizing these inhibitory compounds perform steady-state measurements after prolonged treatment, the results are difficult to assess because the compensatory protein folding stress response has altered the expression of all other proteostasis network genes. Proteasome inhibitors have also been developed with great success, but experimental applications are also limited by unknown off-target effects (which differ between different compounds) and issues concerning a compensatory stress response activating due to widespread misfolding[34-37].

Transcription factor modulators are an interesting class of small molecule for studying proteostasis. Since proteostasis components often perform certain functions while complexed

with other proteostasis network components, stress responsive transcription factors functionally upregulate each of these components together[14, 38-40]. Small molecule activators of proteostasis network transcription factors could be utilized to conveniently upregulate a variety of genes in a manner that is similar to what is observed naturally. Unfortunately, small molecule activators of transcription factors are limited to only a handful of molecules, and most activators typically function indirectly by causing undesirable and toxic stress[41, 42].

One promising method for addressing the challenges associated with altering proteostasis networks through modulation of transcription factors involves a chemical-genetic approach to controlling stress response genes[43-45]. In recent work, investigators have demonstrated that stress-independent activation of proteostasis networks can be achieved through tightly regulated expression of dominant positive transcription factors fused to destabilizing domains. These destabilizing domains (along with any genetically fused protein, such as a transcription factor) are recognized by the ubiquitin-proteasome pathway and degraded, unless a stabilizing small molecule is added. By fusing a dominant positive version of a stress responsive transcription factor, the expression levels of chaperones and quality control factors can be increased simply by adding more of the small molecule. Importantly, persistent overexpression of transcription factors is often cytotoxic, so traditional genetic methods simply will not work for HSF1 and many other proteostasis genes.[43] As a result, the development of the regulated and dosable destabilizing domain-regulated transcription factor approach has enabled a wide variety of experiments which were previously difficult or intractable [46-55].

In principle, inhibition of HSF1 could be achieved in a similar manner using regulation of a dominant negative transcription factor, and this approach could be combined with any of the aforementioned inhibitors of chaperone or quality control components to uncouple compensatory HSF1 activity from chaperone inhibition. Such a method would help clarify whether experimental results can be attributed to the lack of chaperone activity or the activation

of HSF1. However, in the case of HSF1, there is no known small molecule inhibitor or other convenitent method to inhibit HSF1 activity[56-59]. One reason for the lack of a small molecule inhibitor of HSF1 is the limited structural knowledge which is available for HSF1. This lack of structural knowledge comes as no surprise, as HSF1 is thought to be an intrinsically disordered protein[60]. In an effort to create an approach that addresses the apparent deficiency of HSF1 inhibitors, I describe in Chapter 2 a highly transportable method which utilizes a chemical-genetic approach to confer user-defined, highly specific, and small molecule-mediated inhibition of HSF1 activity in any mammalian cell system for the first time[33].

The difficulties associated with the development of modulators of HSF1 serves as an indicator to a notable technical problem concerning traditional pharmacological approaches targeting proteins in a dynamic gene network. Since proteostasis networks and other adaptive stress responses are often comprised of complex gene sets with overlapping functions, small molecule-based inhibitors of individual components may always produce unclear results due to compensatory stress response activation and other pleotropic effects. Investigators can develop specialized methods to address compensatory stress responses as needed, but this approach would require significant engineering in addition to the development of pharmacological modulators, which can take years to develop for even the highest profile targets (like HSF1). Even though great strides have been made to develop methods that modulate the activities of particular components of proteostasis networks (such as chaperone inhibitors, proteasome inhibitors and transcription factor modulators), there are still hundreds of components across various proteostasis networks in separate cellular compartments with no convenient method of modulating activity. Furthermore, traditional small molecule-based perturbation methods may not be a viable method for elucidating the function of many components of proteostasis networks in the near future. In order to address the apparent limitation in traditional pharmacological approaches applied to complex gene networks, I describe in Chapter 3 an

approach to modulate the activities of one (or many) components of a complex gene network by utilizing a small molecule-regulated, dosable version of CRISPR-Cas9 transcription factors to tune expression levels of endogenous genes[61].

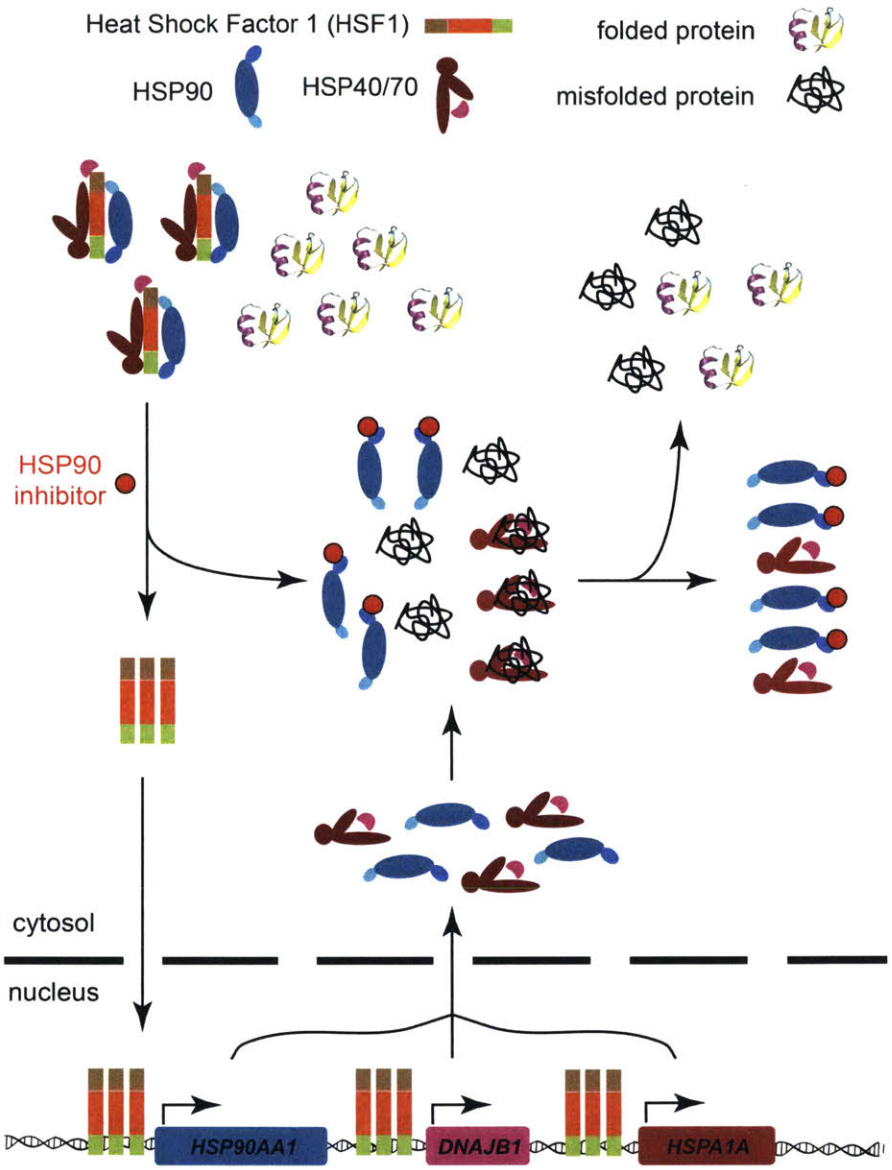

Figure 1.4 | Chaperone inhibition causes protein misfolding and compensatory stress response

Schematic demonstrating the effects of treatment with an HSP90 inhibitor or some other small molecule inhibitor of chaperone activity. Loss of HSP90 activity results in protein misfolding, ultimately causing HSF1 to respond by producing more HSP90 and other cytosolic proteostasis network components. HSP90 remains inhibited, but experimental observations may be attributed to altered activity of other proteostasis network components, an issue which is widely ignored in the literature.

1.2 Evolutionary Approaches to Study Proteostasis

1.2.1 Introduction to Evolution as a Tool to Study Proteostasis Networks

A complimentary approach to study the function of proteostasis network components involves the use of experimental evolution. In one implementation of this approach, genetically modified organisms overexpressing a proteostasis network component were allowed to freely grow without selection pressure (genetically drift) prior to assessment of diversity. Ultimately, this work revealed that chaperones can increase genetic diversity by buffering theoretically destabilizing mutations that would typically cause a protein to have a greater propensity to misfold[62-64]. A separate approach involving the use of HSP90 inhibitors to perturb metazoan proteostasis has demonstrated that proteostasis network remodeling can reveal cryptic phenotypes capable of providing a fitness advantage under certain environmental circumstances[55, 65-70].

Collectively, these experiments have established a connection between proteostasis networks and evolution by demonstrating that chaperones can alter protein evolution. This prior work led us to wonder if the experimental question could be reversed: could molecular evolution be leveraged to study proteostasis networks? By monitoring the evolution of a specific protein in normal and perturbed proteostasis networks, perhaps we could learn more about the functions of specific proteostasis genes and features of the client proteins that engage certain proteostasis systems. In order to test for differences in protein evolution in perturbed proteostasis environments, we need (1) potent and scalable methods that allow us to freely modulate proteostasis networks in living systems, and (2) robust experimental evolution systems that allow us to explore protein sequence space on the laboratory time scale in metazoan cells. While the existing set of tools that perturb proteostasis network components was expanding through the development of chemical-genetic methods in our lab, the methods

for robustly performing experimental evolution in living systems remained extremely limited, particularly in metazoan cells.

1.2.2 Continuous Directed Evolution: A Robust Method for Exploring Sequence Space

Perhaps the most robust experimental evolution methods are utilized for directed evolution, which is a process that aims to discover biomolecules with new and improved activities[71-73]. Directed evolution approaches can be separated into three steps; (1) mutagenesis to create a diverse library, (2) selection to assess for improved activities and (3) amplification to enrich variants possessing improved activity (Figure 1.). In a directed evolution experiment designed to optimize a small region of a protein, all of these steps can be performed *in vitro.* However, directed evolution experiments requiring more mutations across larger, more complex proteins can cause conventional *in vitro* methods to be incredibly laborious to implement, especially in metazoan cells, and ultimately produce poor results.

Moreover, directed evolution experiments aiming to evolve larger proteins with complex activities are intractable with *in vitro* methods because protein sequence space is incredibly massive. Protein sequence space scales as a function of protein length, since there are 20 potential amino acids at each position. Thus, a small protein the size of green fluorescent protein (238 residues) has ~4.42×10^{309} combinatorial possibilities that can be explored, a value that astronomically dwarfs the number of estimated atoms in the observable universe (~10^{80}, calculated from estimated mass of the observable universe[74]) and the Shannon number[75] (theoretical conservative lower bound to potential chess games that can be played, ~10^{120}), multiplied together. The options increase significantly as a protein length approaches the average size of known proteins in the human genome (469 residues, corresponding 1.5×10^{610} combinatorial possibilities)[76]. Further aspects to consider, such as post-translational modifications and the desire to evolve activities involving hetero-multimeric proteins, only increase the size and complexity of protein sequence space.

In an effort to address the problem associated with exploring a sequence space which is so potentially massive, investigators have developed new directed evolution methodologies which seek to identify significant bottlenecks in directed evolution experimental workflows that hinder protein sequence space exploration[77]. In one prominent example, researchers identified that human intervention at different points and time can significantly constrain sequence space exploration[78]. Depending on the model system and screening methods which are used, an *in vitro* directed evolution experiment can require multiple days to perform a single round of diversification, expression and selection. By coupling each of these steps to the viability of a self-replicating organism that quickly reproduces, multiple rounds of directed evolution can be performed in succession on the laboratory time scale. In a notable version of this continuous directed evolution method termed phage-assisted continuous directed evolution (PACE)[78], a phage carrying a biomolecule of interest must replicate quickly to beat out the rate of dilution in a continuously growing culture. By coupling biomolecular activity to the viability of the phage, dozens of rounds of evolution can be achieved *in a single day*. In summary, the amount of time required to complete a single round of evolution is significantly compressed by transitioning steps of the directed evolution process into a living system which runs continuously, ultimately enabling a progressive exploration of a greater amount of sequence space over time.

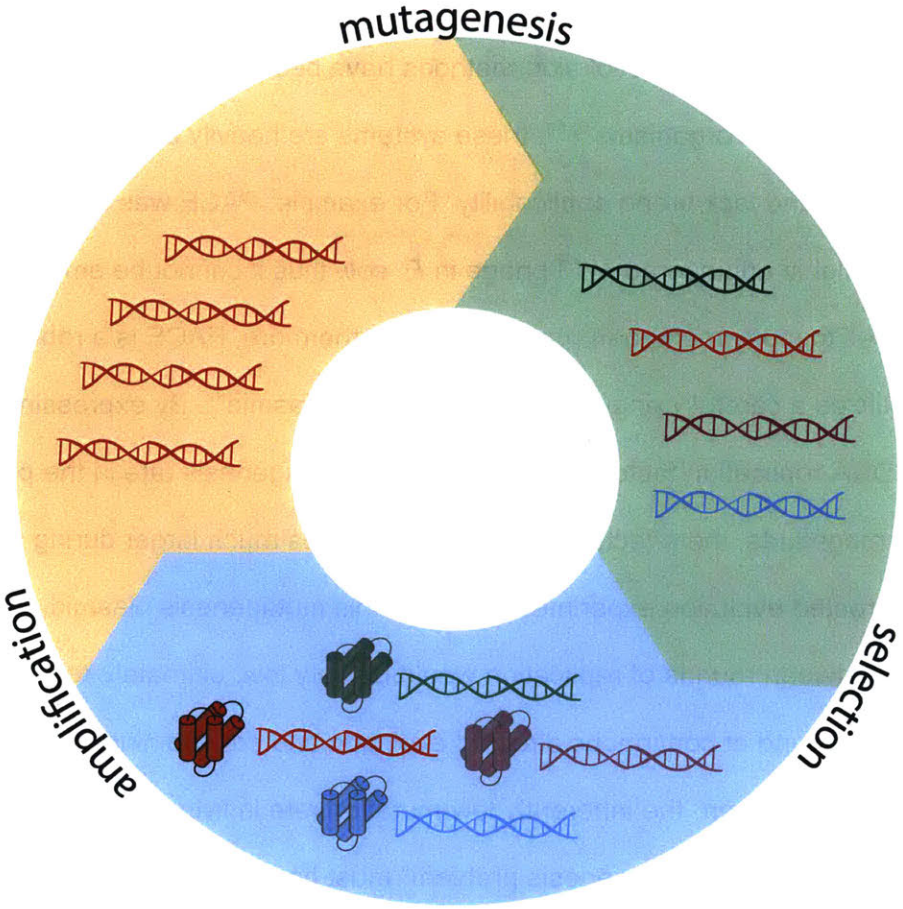

Figure 1.5 | Directed evolution optimizes biomolecular activity

A diagram demonstrating the steps of the directed evolution process. A library of genetic variants is initially made through mutagenesis. Active variants are selected by screening. The library is ultimately enriched by amplifying active variants. Each of these steps can be performed *in vitro*, but living systems are capable of connecting these steps to carry out evolution *in vivo*.

1.2.3 Overcoming the Global Mutagenesis Problem

While continuous directed evolution methods have been established and utilized to great success in some model organisms[79-83], these systems are heavily engineered for particular model organisms and lack broad applicability. For example, PACE was designed to work by propagating a highly engineered M13 phage in *E. coli*, thus it cannot be employed in a mammalian cell to study metazoan proteostasis. Furthermore, PACE is a robust system because it utilizes a carefully engineered mutagenesis plasmid[84]. By expressing a number of error-prone DNA replication factors that increase the mutagenesis rate in the propagating phage by orders of magnitude, the effective library size becomes much larger during the course of the continuous directed evolution experiment. Without this mutagenesis plasmid, the mutation rate of the phage between rounds of replication would be very low, ultimately reducing the number of variants in each round of continuous directed evolution and constraining protein sequence space exploration. In short, the inherently low mutation rate in living systems produces very low genetic diversity, and this "mutagenesis problem" must be overcome in all continuous directed evolution[77, 84, 85].

Initial approaches for increasing mutagenesis rates in living systems relied on chemical and physical mutagens which incorporate mutations globally, or throughout all DNA present in an organism[86-88]. Treatment conditions with these exogenous mutagens requires extensive optimization as higher amounts of mutagenesis eventually causes an accumulation of mutations in essential genes which causes loss of viability (**Figure 1.6**, left side)[87]. Even with optimized treatment conditions, undesired mutations also appear in genes which allow an organism to escape phenotypic selection. In the context of directed evolution, these phenotypic escape variants results in the selection and subsequent enrichment of false positives (**Figure 1.7**, left side)[89]. Either of these types of off-target mutations would be deleterious to evolution experiments designed to study proteostasis as well.

Further development of mutagenesis methods have transitioned to mechanisms that diversify DNA *in vivo* without repetitive optimization between rounds of directed evolution. Living systems can be engineered endogenously express error-prone DNA polymerases and other DNA-damaging enzymes which require minimal intervention during the course of an experiment[84, 90, 91]. Thus, continuous directed evolution methods have opted to utilize plasmids encoding these error-prone DNA replication factors[77]. However, mutagenic plasmids still require significant optimization *prior to an experiment,* as error-prone replication still mutagenizes DNA globally. Unfortunately, the issue of deleterious mutation accumulation still applies as long as global mutagenesis methods are utilized, forcing experimentalists to continuously develop optimized error-prone DNA replication methods for the living system they are studying.

In theory, a targeted mutagenesis method would minimize off-target mutations in essential genes, resulting in fewer false negatives in an evolution experiment (**Figure 1.6**, right side). Targeted mutagenesis would also reduce false positives originating from undesired off-target mutations that allow cheating of a phenotypic selection (**Figure 1.7**, right side). Recognizing that off-target mutations caused by global mutagenesis hinders the application of continuous directed evolution to study proteostasis, we set out to develop a generalized targeted mutagenesis method described in Chapter 4, which will enable investigators to perform more robust, *in vivo* continuous evolution experiments.

By developing a targeted mutagenesis method that localizes mutations to an individual protein, we envision that evolution experiments studying the function of proteostasis network components will become more tractable and informative. Investigators will have the ability to study how a protein evolves in different cellular protein folding environments and learn which type of mutations are accessible or deleterious. By understanding how these evolutions are stabilized or preferentially removed, these experiments may ultimately inform how particular components of the proteostasis network function to alter protein folding in living systems.

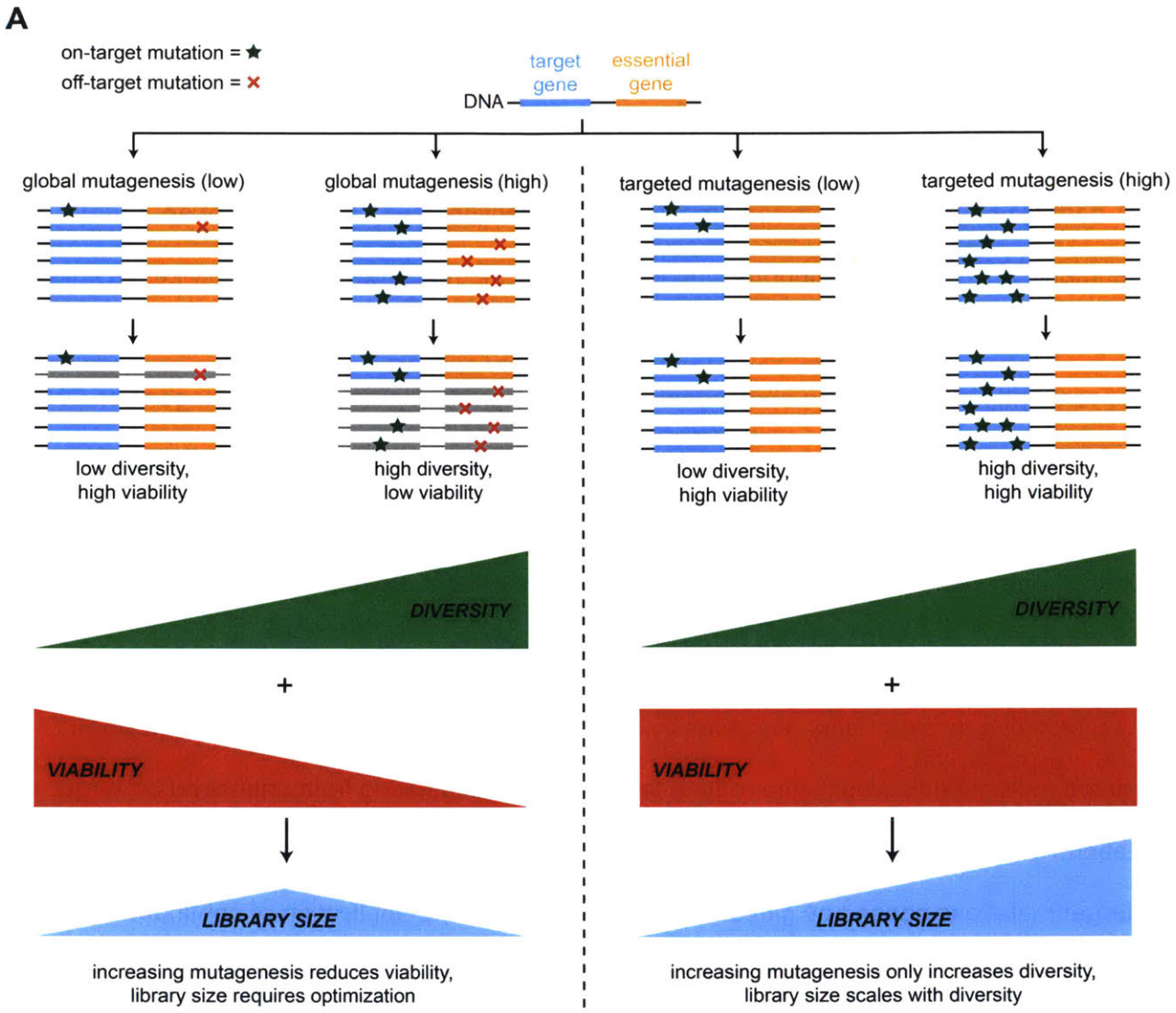

Figure 1.6 | Global mutagenesis causes deleterious mutations in essential genes

Global mutagenesis indiscriminately introduces mutations across an organism's entire genome, introducing mutations in target genes and other genes such as essential genes. Attempts to increase the global mutagenesis rate and thus library diversity lead to decreased cell viability due to off-target mutations in essential genes. Targeted mutagenesis allows for a high mutagenesis rate that does not decrease cell viability by minimizing off-target mutations in essential genes.

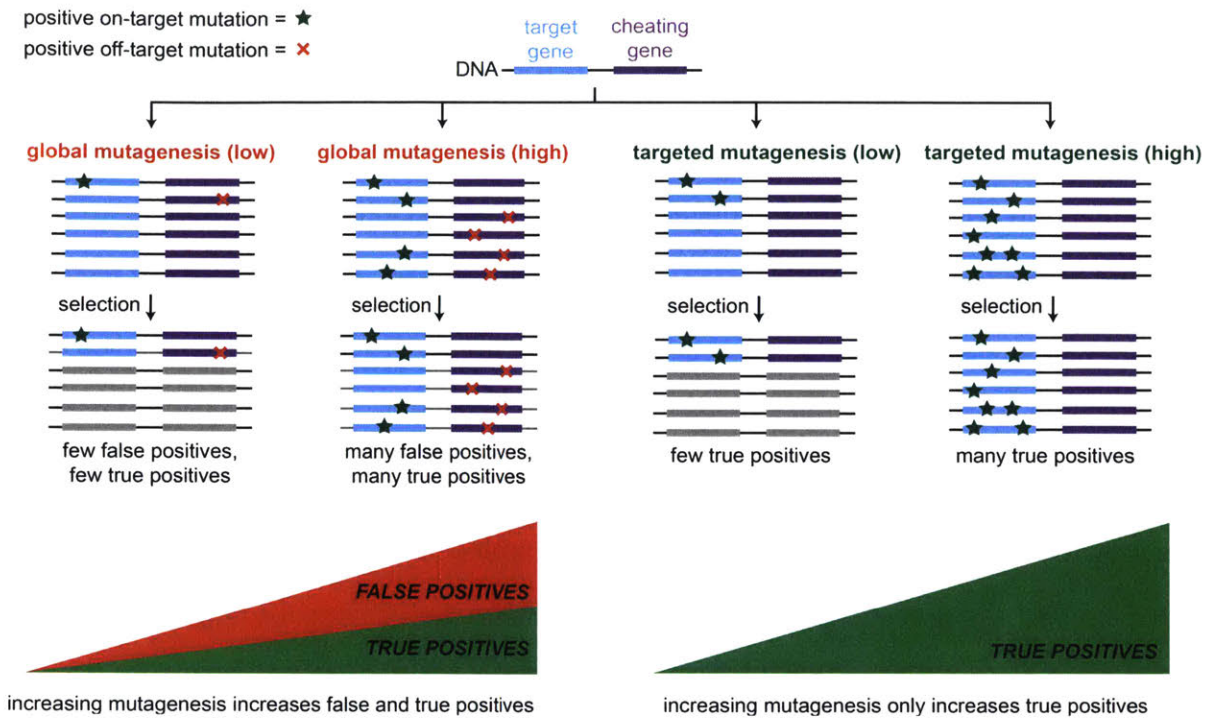

Figure 1.7 | Off-target mutations promotes cheating of selection

Global mutagenesis incorporates off-target mutations in genes that allow an organism to cheat a phenotypic selection. In an evolution experiment, this will cause a certain rate of false positives. Targeted mutagenesis minimizes these false positives by preventing off-target mutations in genes that allow the organism to cheat the selection.

1.3 Concluding Remarks

In summary, the study of proteostasis networks is technically difficult because these gene networks are composed of multiple components that work together to mediate the protein folding process in living systems. The development of improved inhibitory molecules, specialized methods which modulate transcription factor activities, and tunable Cas9-based genetic perturbations now enables investigators to perform previously intractable studies and will lead to the discovery of further functions for many components of proteostasis networks function. Furthermore, the development of targeted mutagenesis methods has opened doors by enabling more elegant studies concerning the effects proteostasis network components have on protein evolution. Collectively, these methods demonstrate the tremendous impact chemical biology approaches can have when applied to difficult problems in science.

1.4 References

1. Fraenkel, A.S. Complexity of protein folding. *Bulletin of Mathematical Biology* **55**, 1199-1210 (1993).
2. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195 (1995).
3. Schubert, U. et al. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770-774 (2000).
4. Horowitz, S. et al. Visualizing chaperone-assisted protein folding. *Nature structural & molecular biology* **23**, 691-697 (2016).
5. Brockwell, D.J. & Radford, S.E. Intermediates: ubiquitous species on folding energy landscapes? *Current opinion in structural biology* **17**, 30-37 (2007).
6. Martin, J. & Hartl, F.-U. The effect of macromolecular crowding on chaperonin-mediated protein folding. *Proceedings of the National Academy of Sciences* **94**, 1107-1112 (1997).
7. Balch, W.E., Morimoto, R.I., Dillin, A. & Kelly, J.W. Adapting proteostasis for disease intervention. *Science* **319**, 916-919 (2008).
8. Balchin, D., Hayer-Hartl, M. & Hartl, F.U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).
9. Georgopoulos, C. The emergence of the chaperone machines. *Trends in biochemical sciences* **17**, 295-299 (1992).
10. Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nature Rev. Mol. Cell Biol.* **14**, 630-642 (2013).
11. Peters, J.M., Franke, W.W. & Kleinschmidt, J.A. Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. *Journal of Biological Chemistry* **269**, 7709-7718 (1994).
12. Ravid, T. & Hochstrasser, M. Diversity of degradation signals in the ubiquitin-proteasome system. *Nature reviews. Molecular cell biology* **9**, 679-690 (2008).
13. Åkerfelt, M., Morimoto, R.I. & Sistonen, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell. Biol.* **11**, 545-555 (2010).
14. Voellmy, R. On mechanisms that control heat shock transcription factor activity in metazoan cells. *Cell Stress Chaper.* **9**, 122-133 (2004).
15. Moran Luengo, T., Kityk, R., Mayer, M.P. & Rudiger, S.G.D. Breaking the Deadlock of Molecular Chaperones. *bioRxiv* (2017).
16. Kampinga, H.H. & Craig, E.A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* **11**, 579-592 (2010).
17. Sala, A.J., Bott, L.C. & Morimoto, R.I. Shaping proteostasis at the cellular, tissue, and organismal level. *The Journal of Cell Biology* (2017).
18. Heldens, L. et al. Protein refolding in peroxisomes is dependent upon an HSF1-regulated function. *Cell Stress Chaper.* **17**, 603-613 (2012).
19. Basha, W. et al. Geldanamycin, a potent and specific inhibitor of Hsp90, inhibits gene expression and replication of human cytomegalovirus. *Antivir. Chem. Chemother.* **16**, 135-146 (2005).

20. Burlison, J.A., Neckers, L., Smith, A.B., Maxwell, A. & Blagg, B.S.J. Novobiocin: Redesigning a DNA gyrase inhibitor for selective inhibition of Hsp90. *J. Am. Chem. Soc.* **128**, 15529-15536 (2006).
21. Dymock, B.W. et al. Novel, potent small-molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design. *J Med Chem* **48**, 4212-4215 (2005).
22. Neckers, L. & Workman, P. Hsp90 molecular chaperone inhibitors: are we there yet? *Clin. Cancer Res.* **18**, 64-76 (2012).
23. Sharp, S. & Workman, P. Inhibitors of the HSP90 molecular chaperone: current status. *Adv Cancer Res* **95**, 323-348 (2006).
24. Ying, W.W. et al. Ganetespib, a unique triazolone-containing HSP90 inhibitor, exhibits potent antitumor activity and a superior safety profile for cancer therapy. *Mol. Cancer Ther.* **11**, 475-484 (2012).
25. Garg, G., Zhao, H.P. & Blagg, B.S.J. Design, synthesis, and biological evaluation of ring-constrained Novobiocin analogues as Hsp90 C-terminal inhibitors. *ACS Med. Chem. Lett.* **6**, 204-209 (2015).
26. Khandelwal, A. et al. Structure-guided design of an Hsp90beta N-terminal isoform-selective inhibitor. *Nature communications* **9**, 425 (2018).
27. T. Gewirth, D. Paralog Specific Hsp90 Inhibitors - A Brief History and a Bright Future. *Current Topics in Medicinal Chemistry* **16**, 2779-2791 (2016).
28. Evans, C.G., Chang, L. & Gestwicki, J.E. Heat shock protein 70 (Hsp70) as an emerging drug target. *J. Med. Chem.* **53**, 4585-4602 (2010).
29. Li, X. et al. Analogues of the Allosteric Heat Shock Protein 70 (Hsp70) Inhibitor, MKT-077, As Anti-Cancer Agents. *ACS Medicinal Chemistry Letters* **4**, 1042-1047 (2013).
30. Patury, S., Miyata, Y. & Gestwicki, J.E. Pharmacological targeting of the Hsp70 chaperone. *Current topics in medicinal chemistry* **9**, 1337-1351 (2009).
31. Pratt, W.B., Gestwicki, J.E., Osawa, Y. & Lieberman, A.P. Targeting Hsp90/Hsp70-based protein quality control for treatment of adult onset neurodegenerative diseases. *Annu. Rev. Pharmacol. Toxicol.* (2014).
32. Taylor, I.R. et al. High Throughput Screen for Inhibitors of Protein-Protein Interactions in a Reconstituted Heat Shock Protein 70 (Hsp70) Complex. *Journal of Biological Chemistry* (2018).
33. Moore, C.L. et al. Transportable, chemical genetic methodology for the small molecule-mediated inhibition of heat shock factor 1. to be submitted (2015).
34. Cavo, M. Proteasome inhibitor bortezomib for the treatment of multiple myeloma. *Leukemia* **20**, 1341-1352 (2006).
35. van der Linden, W.A. et al. Discovery of a potent and highly beta 1 specific proteasome inhibitor from a focused library of urea-containing peptide vinyl sulfones and peptide epoxyketones. *Organic & Biomolecular Chemistry* **10**, 181-194 (2012).
36. Crawford, L.J., Walker, B. & Irvine, A.E. Proteasome inhibitors in cancer therapy. *Journal of cell communication and signaling* **5**, 101-110 (2011).
37. Kisselev, A.F., van der Linden, W.A. & Overkleeft, H.S. Proteasome inhibitors: an expanding army attacking a unique target. *Chemistry & biology* **19**, 99-115 (2012).
38. Anckar, J. & Sistonen, L. Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu. Rev. Biochem.* **80**, 1089-1115 (2011).

39. Khalil, S., Luciano, J., Chen, W.J. & Liu, A.Y.C. Dynamic regulation and involvement of the heat shock transcriptional response in arsenic carcinogenesis. *J. Cell. Physiol.* **207**, 562-569 (2006).
40. Zuo, J.R., Rungger, D. & Voellmy, R. Multiple layers of regulation of human heat-shock transcription factor-1. *Mol. Cell. Biol.* **15**, 4319-4330 (1995).
41. West, J.D., Wang, Y. & Morano, K.A. Small molecule activators of the heat shock response: chemical properties, molecular targets, and therapeutic promise. *Chemical research in toxicology* **25**, 2036-2053 (2012).
42. Westerheide, S.D. et al. Celastrols as inducers of the heat shock response and cytoprotection. *J. Biol. Chem.* **279**, 56053-56060 (2004).
43. Ryno, L.M. et al. Characterizing the altered cellular proteome induced by the stress-independent activation of heat shock factor I. *ACS Chem. Biol.*, 1273-1283 (2014).
44. Shoulders, M.D., Ryno, L.M., Cooley, C.B., Kelly, J.W. & Wiseman, R.L. Broadly applicable methodology for the rapid and dosable small molecule-mediated regulation of transcription factors in human cells. *J. Am. Chem. Soc.* **135**, 8129-8132 (2013).
45. Shoulders, M.D. et al. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **3**, 1279-1292 (2013).
46. Chen, J.J. et al. ATF6 activation reduces the secretion and extracellular aggregation of destabilized variants of an amyloidogenic protein. *Chem Biol* **21**, 1564-1574 (2014).
47. Cooley, C.B. et al. Unfolded protein response activation reduces secretion and extracellular aggregation of amyloidogenic immunoglobulin light chain. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13046-13051 (2014).
48. Dewal, M.B. et al. XBP1s Links the Unfolded Protein Response to the Molecular Architecture of Mature N-Glycans. *Chemistry & biology* **22**, 1301-1312 (2015).
49. Genereux, J.C. et al. Unfolded protein response-induced ERdj3 secretion links ER stress to extracellular proteostasis. *The EMBO journal* **34**, 4-19 (2015).
50. Plate, L. & Cooley, C.B. Small molecule proteostasis regulators that reprogram the ER to reduce extracellular protein aggregation. **5** (2016).
51. Ryno, L.M. et al. Characterizing the Altered Cellular Proteome Induced by the Stress-Independent Activation of Heat Shock Factor 1. *ACS Chemical Biology* **9**, 1273-1283 (2014).
52. Wang, X. et al. The Systemic Amyloid Precursor Transthyretin (TTR) Behaves as a Neuronal Stress Protein Regulated by HSF1 in SH-SY5Y Human Neuroblastoma Cells and APP23 Alzheimer's Disease Model Mice. *The Journal of Neuroscience* **34**, 7253-7265 (2014).
53. Plate, L. et al. Small molecule proteostasis regulators that reprogram the ER to reduce extracellular protein aggregation. *eLife* **5** (2016).
54. Shoulders, M.D. et al. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell reports* **3**, 1279-1292 (2013).
55. Phillips, A.M. et al. Host proteostasis modulates influenza evolution. *eLife* **6** (2017).
56. Whitesell, L. & Lindquist, S. Inhibiting the transcription factor HSF1 as an anticancer strategy. *Exp. Op. Ther. Targets* **13**, 469-478 (2009).

57. Au, Q.Y., Zhang, Y.J., Barber, J.R., Ng, S.C. & Zhang, B. Identification of inhibitors of HSF1 functional activity by high-content target-based screening. *J. Biomol. Screen.* **14**, 1165-1175 (2009).
58. Santagata, S. et al. Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. *Science* **341**, 1238303 (2013).
59. Westerheide, S.D., Kawahara, T.L.A., Orton, K. & Morimoto, R.I. Triptolide, an inhibitor of the human heat shock response that enhances stress-induced cell death. *J. Biol. Chem.* **281**, 9616-9622 (2006).
60. Westerheide, S.D., Raynes, R., Powell, C., Xue, B. & Uversky, V.N. HSF transcription factor family, heat shock response, and protein intrinsic disorder. *Current protein & peptide science* **13**, 86-103 (2012).
61. Maji, B. et al. Multidimensional chemical control of CRISPR-Cas9. *Nature chemical biology* **13**, 9-11 (2017).
62. Tokuriki, N. & Tawfik, D.S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668-671 (2009).
63. Tokuriki, N. & Tawfik, D.S. Stability effects of mutations and protein evolvability. *Curr. Op. Struct. Biol.* **19**, 596-604 (2009).
64. Williams, T.A. & Fares, M.A. The effect of chaperonin buffering on protein evolution. *Genome Biol. Evol.* **2**, 609-619 (2010).
65. Queitsch, C., Sangster, T.A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618-624 (2002).
66. Rohner, N. et al. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. *Science* **342**, 1372-1375 (2013).
67. Rutherford, S.L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336-342 (1998).
68. Sangster, T.A. et al. Phenotypic diversity and altered environmental plasticity in Arabidopsis thaliana with reduced Hsp90 levels. *PLoS One* **2**, e648 (2007).
69. Sangster, T.A. et al. HSP90-buffered genetic variation is common in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA* **105**, 2969-2974 (2008).
70. Sangster, T.A. et al. HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. *Proc. Natl. Acad. Sci. USA* **105**, 2963-2968 (2008).
71. Arnold, F.H. Design by directed evolution. *Acc. Chem. Res.* **31**, 125-131 (1998).
72. Romero, P.A. & Arnold, F.H. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10**, 866-876 (2009).
73. Packer, M.S. & Liu, D.R. Methods for the directed evolution of proteins. *Nature reviews. Genetics* **16**, 379-394 (2015).
74. Davies, P. The Goldilocks Enigma: Why Is the Universe Just Right for Life? (Mariner Books, 2008).
75. Shannon, C.E. Programming a Computer for Playing Chess. 2-13 (1988).
76. International Human Genome Sequencing, C. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).
77. Badran, A.H. & Liu, D.R. In vivo continuous directed evolution. *Curr. Opin. Chem. Biol.* **24**, 1-10 (2014).

78. Esvelt, K.M., Carlson, J.C. & Liu, D.R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503 (2011).
79. Dickinson, B.C., Packer, M.S., Badran, A.H. & Liu, D.R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nature Comm.* **5**, 6352 (2014).
80. Carlson, J.C., Badran, A.H., Guggiana-Nilo, D.A. & Liu, D.R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chem. Biol.* **10**, 216-222 (2014).
81. Hubbard, B.P. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature methods* **12**, 939-942 (2015).
82. Badran, A.H. et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **533**, 58-63 (2016).
83. Bryson, D.I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nature chemical biology* **13**, 1253 (2017).
84. Badran, A.H. & Liu, D.R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
85. Gago, S., Elena, S.F., Flores, R. & Sanjuan, R. Extremely high mutation rate of a hammerhead viroid. *Science* **323**, 1308-1308 (2009).
86. Hong, J.S. & Ames, B.N. Localized mutagenesis of any specific small region of the bacterial chromosome. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 3158-3162 (1971).
87. Tessman, I., Ishiwa, H. & Kumar, S. Mutagenic Effects of Hydroxylamine in vivo. *Science* **148**, 507-508 (1965).
88. Cupples, C.G. & Miller, J.H. A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. *Proc. Natl. Acad. Sci.* **86**, 5345-5349 (1989).
89. Tizei, P.A., Csibra, E., Torres, L. & Pinheiro, V.B. Selection platforms for directed evolution in synthetic biology. *Biochem. Soc. Trans.* **44**, 1165-1175 (2016).
90. Camps, M., Naukkarinen, J., Johnson, B.P. & Loeb, L.A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9727-9732 (2003).
91. Greener, A., Callahan, M. & Jerpseth, B. An efficient random mutagenesis technique using an E. coli mutator strain. *Mol. Biotechnol.* **7**, 189-195 (1997).

Chapter 2: A Transportable, Chemical Genetic Methodology for the Small Molecule-Mediated Inhibition of Heat Shock Factor 1

This chapter is adapted from the following manuscript:
Moore, C.L., Dewal, M.B., Nekongo, E.E., Santiago, S., Lu, N.B., Levine, S.S., Shoulders, M.D. *ACS Chem. Biol.*, 2016, *11*(1), pp 200-210.

2.1 Author Contributions

C.L.M. clones all constructs with assistance from S.S. and N.B.L. C.L.M. performed and analyzed data for all qPCR experiments in HEK cells. M.B.D. performed Western blot experiments with assistance from C.L.M. E.E.K. generated data for MDCK cells with assistance from C.L.M. C.L.M. performed luciferase and PULSA experiments. S.S.L. oversaw and analyzed data from RNA microarray experiments. M.D.S. supervised all experimentation and data analysis.

2.2 Abstract

Proteostasis in the cytosol is governed by the heat shock response. The master regulator of the heat shock response, heat shock factor 1 (HSF1), and key chaperones whose levels are HSF1-regulated have emerged as high-profile targets for therapeutic applications ranging from protein misfolding-related disorders to cancer. Nonetheless, a generally applicable methodology to selectively and potently inhibit endogenous HSF1 in a small molecule-dependent manner in disease model systems remains elusive. Also problematic, the administration of even highly selective chaperone inhibitors often has the side effect of activating HSF1 and thereby inducing a compensatory heat shock response. Herein, we report a ligand-regulatable, dominant negative version of HSF1 that addresses these issues. Our approach, which required engineering a new dominant negative HSF1 variant, permits dosable inhibition of endogenous HSF1 with a selective small molecule in cell-based model systems of interest. The methodology allows us to uncouple the pleiotropic effects of chaperone inhibitors and environmental toxins from the concomitantly induced compensatory heat shock response. Integration of our method with techniques to activate HSF1 enables the creation of cell lines in which the cytosolic proteostasis network can be up- or down-regulated by orthogonal small molecules. Selective, small molecule-mediated inhibition of HSF1 has distinctive implications for the proteostasis of both chaperone-dependent globular proteins and aggregation-prone intrinsically disordered

proteins. Altogether, this work provides critical methods for continued exploration of the biological roles of HSF1 and the therapeutic potential of heat shock response modulation.

2.3 Introduction

The heat shock response (HSR) maintains cytosolic proteostasis by dynamically matching cellular protein folding capacity to demand and environmental conditions.[1] Acute protein folding insults, such as heat or oxidative stress, as well as chronic misfolding and protein aggregation activate the HSR, thereby transcriptionally upregulating chaperones and quality control factors that coordinate to address the proteostatic challenge.[2] Induction of these proteostasis mechanisms is mediated by the master regulator of the HSR, the transcription factor heat shock factor 1 (HSF1).[3] Under unstressed conditions, the majority of cellular HSF1 is maintained in an inactive state. Protein misfolding insults instigate a shift toward nuclear, trimeric, and transcriptionally active HSF1.[4] Molecular details of the HSF1 activation mechanism remain under investigation, but chaperone binding to maintain HSF1 in its inactive, monomeric form and post-translational modifications play an important role.[5, 6]

HSF1 inhibition has recently emerged as a promising therapeutic strategy for a diverse array of cancers.[7] Of particular note, genetic ablation of HSF1 attenuates resistance to tumorigenesis in mice,[8] constitutive upregulation of HSF1 is linked to poor breast cancer patient prognosis,[9] and HSF1 activation in cancer-associated fibroblasts promotes malignancy.[10] Interestingly, HSF1 in malignant cells regulates the expression of numerous genes not typically associated with the HSR,[11] suggesting multiple roles for this master regulator of proteostasis in pathologic processes. These and other findings motivate HSF1 inhibition as a potentially useful anticancer therapeutic strategy.[12]

Surprisingly, HSF1 inhibition is also emerging as a possible strategy for diseases linked directly to protein misfolding or aggregation.[13] Balch and co-workers identified a "maladapted stress response" mediated by chronic upregulation of HSF1 in diverse protein misfolding-related disorders, including cystic fibrosis.[14] They found that genetic elimination of HSF1 restored disease-associated proteostasis defects, possibly by reducing hyperactive quality control mechanisms chronically

upregulated by HSF1. The counterintuitive possibility that HSF1 inhibition could be valuable in certain protein misfolding diseases further enhances the appeal of small molecule-based methods to explore the therapeutic potential of HSF1 inhibition.

These and other findings have engendered active efforts to discover inhibitors of HSF1.[7, 15] Unfortunately, the identification of potent and selective small molecule-based inhibitors of transcription factors remains a challenging endeavor. Although a number of compounds capable of inhibiting HSF1 have been identified, including triptolide, NZ28, rohitinib, cantharidin, and more,[7, 15-18] mechanistic studies show that these compounds have pleiotropic effects at the concentrations required for HSF1 inhibition, and several lack potency.[16, 19, 20] In the absence of selective and potent small molecule inhibitors of HSF1, researchers have focused on genetic techniques to ablate HSF1 activity, such as RNAi against HSF1 or overexpression of dominant negative HSF1 variants that can inhibit the endogenously expressed transcription factor.[14, 21-26] These approaches have the advantage of high selectivity and potency, but at the cost of slow kinetics, poor dosability, and cellular adaptation or even death upon chronic HSF1 knockdown or constitutive dominant negative HSF1 overexpression.[27]

An alternative to HSF1 inhibition is targeting the activities of downstream effectors of proteostasis. In this regard, inhibitors of the cytosolic HSP90 and HSP70 chaperones are of substantial interest. Clinical trials testing HSP90 inhibitors in oncology are ongoing, and increasingly potent and selective HSP70 inhibitors continue to be developed.[28, 29] Such chaperone inhibitors are ubiquitously deployed in the literature to evaluate effects of chaperone inhibition on client protein folding and function. A sometimes underappreciated complication for mechanistic studies does, however, exist: many chaperone inhibitors activate a compensatory HSR mediated by HSF1.[30] The consequence is that chaperone inhibition can actually induce high levels of other chaperones and quality control mechanisms, potentially engendering the wrongful attribution of observed phenotypes directly to chaperone inhibition rather than to the compensatory HSR. Similarly problematic, the compensatory HSR can obscure actual consequences of chaperone inhibition.[31] In malignant cells,

this issue can sometimes be resolved by very carefully fine-tuning the dose of a HSP90 inhibitor to sidestep the compensatory HSR, but the acceptable dose range for an inhibitory effect without HSR induction is narrow and often nonexistent. Alternatively, a promising new class of HSP90 inhibitor is emerging that functions by binding to the chaperone's C-terminus and does not activate the HSR.[32, 33] These compounds are under continued development for potency and selectivity.[34] Thus, methods to uncouple the direct effects of chaperone inhibition from the compensatory HSR would be very valuable, not just for elucidating the mechanistic consequences of chaperone inhibition but also for studying the roles of the HSR in protection against other HSR activators like the environmental toxin arsenite.[35]

Small molecule-based methods to potently and selectively inhibit HSF1 are, therefore, urgently required. Here, we report the development of new small molecule-regulated, genetically encoded dominant negative versions of HSF1 that can be easily deployed in diverse model systems of interest. Our approach enables the robust uncoupling of the HSF1-activating side effects of chaperone inhibition or oxidative stress from their direct effects on cellular proteostasis. Importantly, our method is synergistic with and orthogonal to a previously reported approach for small molecule-regulated *activation* of HSF1,[36] allowing us to inducibly activate or repress the cytosolic proteostasis network with small molecules in a single cell, as desired. Finally, we evaluate the implications of small molecule-mediated HSF1 inhibition for the proteostasis of model globular and aggregating cytosolic chaperone clients. Altogether, our work provides a robust methodology valuable for continued studies of the normal and pathologic roles of HSF1, and will inform the continued development of HSF1 regulators for applications in cancer and protein misfolding-related diseases.

2.4 Results and Discussion

2.4.1 Engineering a Potent Dominant Negative Version of Constitutively Active HSF1

Our first objective was to leverage destabilized domain (DD) technology to generate a small molecule-regulated dominant negative version of HSF1 based on extant dominant negative variants.[22, 25, 26] DD fusion suppresses the cellular levels of fusion proteins because the small

DD degron rapidly directs the fusion protein to the proteasome for degradation. Administration of a small molecule that stabilizes the DD prevents degradation and allows the fusion protein to function.[37-39] Transcription factors can be fused to DDs to permit small molecule-dependent, highly dosable induction of transcription factor activity.[36, 40] The methodology is readily transportable, demanding minimal optimization and requiring the introduction of only a single genetic construct to bestow small molecule dose-dependent regulation of transcription factor activity.

Current dominant negative HSF1 variants typically involve deletion of a significant fraction of the C-terminal transcription activation domain of HSF1 (amino acids 379–529).[22, 25, 26] Our early efforts linking such dominant negative constructs to DDs indicated modest potency, suggesting that re-engineering the dominant negative HSF1 protein would be beneficial. A constitutively active version of HSF1, termed cHSF1, in which a portion of the internal regulatory domain of HSF1 (amino acids 186–202) is deleted, was previously characterized.[22, 41] Induction of cHSF1 results in constitutive upregulation of HSF1-dependent genes, even in the absence of HSR activation. We rationalized that a dominant negative version of this cHSF1 variant in which the transcription activation domain (amino acids 379–529) is also deleted would prove to be a highly potent HSF1 inhibitor. Such a construct would not be subject to endogenous mechanisms for regulating HSF1 that constitutively maintain the transcription factor in its inactive, monomeric state[6] and thereby potentially reduce the potency of previously described dominant negative HSF1 variants (**Figure 2.1**). We termed this new dominant negative version of HSF1, lacking both a portion of the internal regulatory domain and the transcription activation domain, "dn-cHSF1."

We first assessed whether our new dn-cHSF1 construct is capable of inhibiting HSF1 by creating stable HEK293T-REx cell lines expressing dn-cHSF1 under control of the doxycycline (dox)-dependent tetracycline repressor. We predicted that induction of dn-cHSF1 would

abrogate the ability of arsenite or HSP90 inhibitor-mediated activation of endogenous HSF1 to upregulate established HSR target genes. As expected, treatment of cells with arsenite or the HSP90 inhibitor STA-9090[42] results in robust induction of the HSR-regulated chaperones *HSP90 (HSP90AA1)*, *HSP70 (HSPA1A)*, and *HSP40 (DNAJB1)*. However, pretreatment with dox for 18 h to induce dn-cHSF1 completely inhibits the arsenite- and STA-9090-mediated upregulation of these HSR genes, as observed by both RT-qPCR and Western blotting (**Figure 2.2** and **Figure 2.3**, primers for **Figure 2.2** in **Table 2.1**). Thus, our new dn-cHSF1 construct is highly effective at inhibiting the HSR-mediated activation of endogenous HSF1.

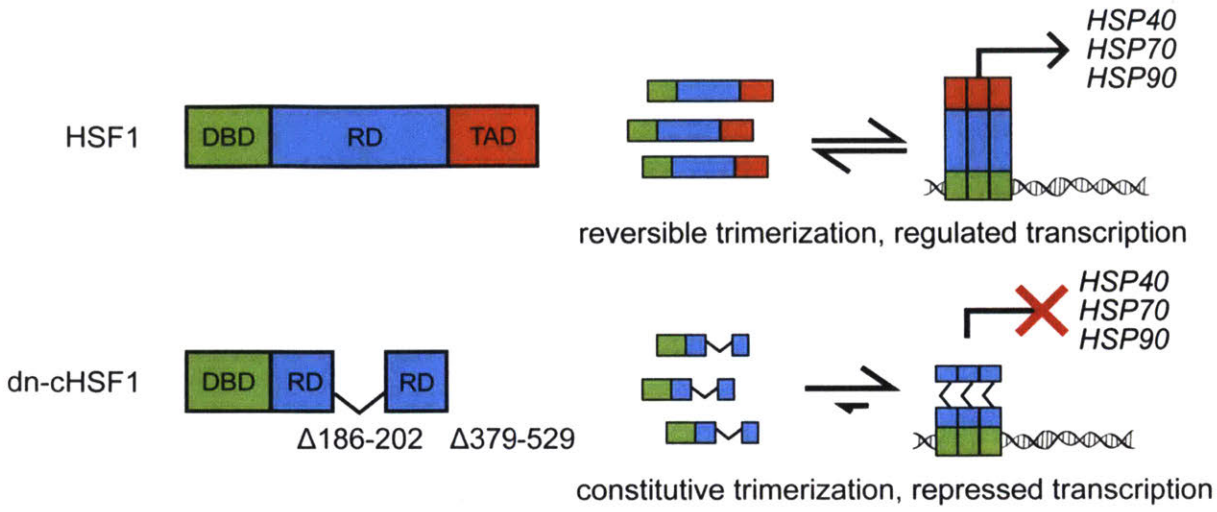

Figure 2.1 | Design of a new, potent dominant negative HSF1 variant "dn-cHSF1"

Model of dn-cHSF1, which lacks a portion of the internal regulatory domain (RD; deleted amino acids 186–202) and the C-terminal transcription activation domain (amino acids 379–529).

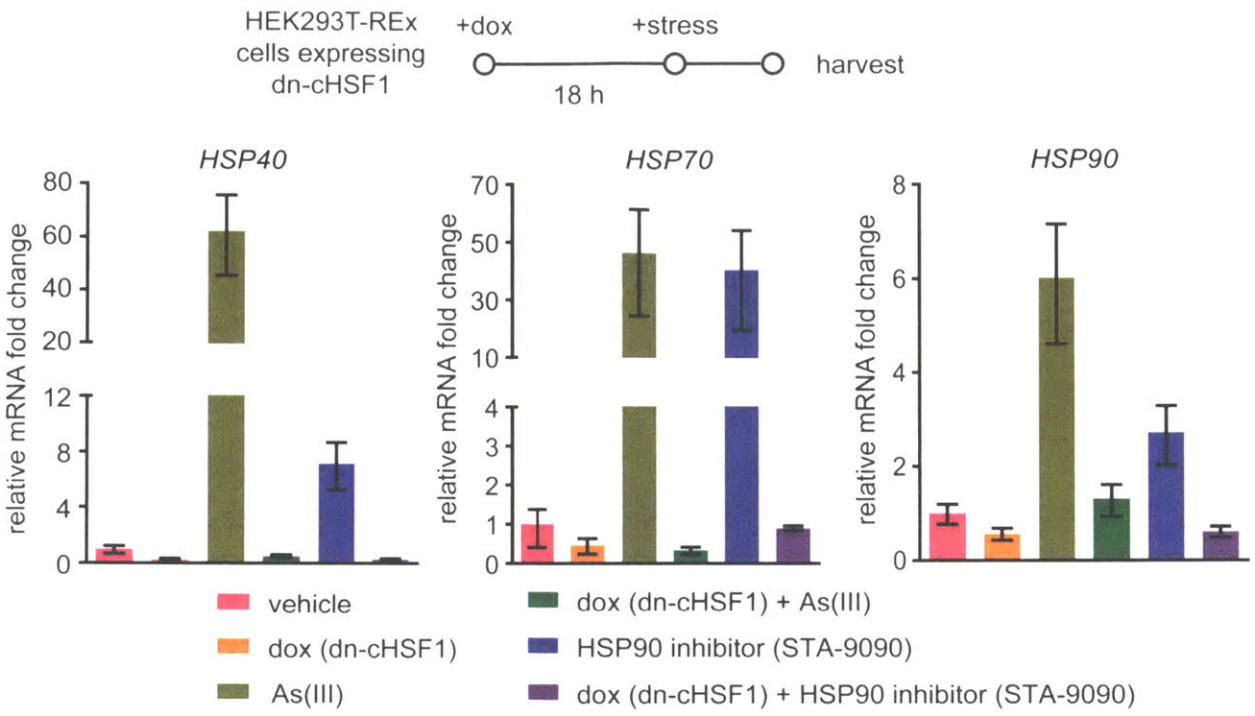

Figure 2.2 | Functional analysis of new dominant negative HSF1 construct

qPCR analysis of *HSP40 (DNAJB1)*, *HSP70 (HSPA1A)*, and *HSP90* (*HSP90AA1*) in HEK293T-REx cells inducibly expressing dn-cHSF1 following pretreatment with vehicle or dox (18 h; 1 µg/mL) and then HSR activation by treatment with arsenite (6 h; 100 µM) or STA-9090 (6 h; 100 nM). qPCR data are reported as the mean ± 95% confidence interval relative to vehicle-treated HEK293T-REx cells.

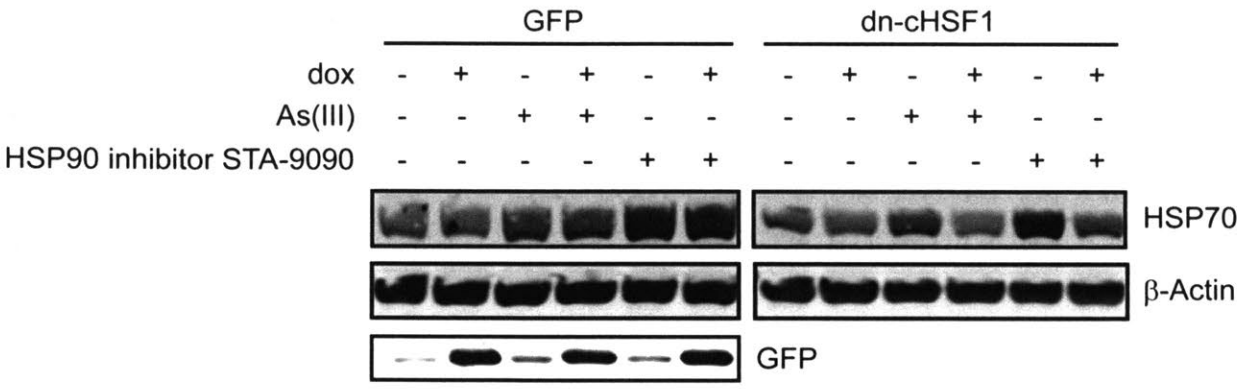

Figure 2.3 | New dominant negative HSF1 prevents chaperone protein level upregulation during heat stress recovery

Immunoblot of HEK293T-REx cells inducibly expressing dn-cHSF1 following pretreatment with vehicle or dox (18 h; 1 μg/mL) and then HSR activation by treatment with arsenite (6 h; 100 μM) or STA-9090 (6 h; 100 nM)

Table 2.1 | Compilation of primers for qPCR used in Figure 2.2

Target gene	Forward primer	Reverse primer
HSP90AA1	GATAAACCCTGACCATTCC	AAGACAGGAGCGCAGTTTCATAAA
HSPA1A	GGAGGCGGAGAAGTACA	GCTGATGATGGGGTTACA
DNAJB1	TGTGTGGCTGCACAGTGAAC	ACGTTTCTCGGGTGTTTTGG
Rplp2	CCATTCAGCTCACTGATAACCTTG	CGTCGCCTCCTACCTGCT

2.4.2 dn-cHSF1 Represses HSR-Mediated Upregulation of HSF1 Target Genes with High Selectivity

We next employed whole genome transcript arrays to assess changes in the transcriptome of cells inducibly expressing dn-cHSF1 both before and after stressing with arsenite or STA-9090. Arsenite is an HSR activator that also has pleiotropic effects including oxidative stress and DNA damage. STA-9090 is a widely employed HSP90 inhibitor that rapidly activates a compensatory HSR. dn-cHSF1 provides a potential mechanism to uncouple the HSR-activating insults of arsenite or STA-9090 from the other consequences of exposure. K-Means clustering of the 3167 genes that had a significant change in mRNA levels under any condition compared to the vehicle (see **Table 2.2** for selected genes from the comprehensive array data) identified a single cluster that displays significant downregulation in response to activation of dn-cHSF1 by dox treatment (**Figure 2.4**, top cluster). Gene ontology analysis revealed that this set of genes is highly enriched for HSR genes. In contrast, the large number of other genes activated by arsenite treatment (including zinc metalloproteins and genes involved in oxidative stress) are not inhibited by dn-cHSF1. Most importantly, only seven genes show a statistically significant reduction of arsenite-induced upregulation owing to dox pretreatment, all of which are classical HSR genes (shown in **Table 2.2** in the cytosolic chaperones group), highlighting the exquisite selectivity of our new dn-cHSF1 construct for HSF1 inhibition.

We next postulated that dn-cHSF1 activation provides a mechanism to uncouple the compensatory heat shock response induced by HSP90 inhibition from the direct consequences of HSP90 inhibition. To test this hypothesis, we treated cells with STA-9090 at the lowest concentration (100 nM) sufficient to inhibit HSP90, as evaluated by the resultant degradation of Akt, an established HSP90 client (**Figure 2.5**).[43] Treatment with 100 nM STA-9090 in our cells robustly induces the HSR, as observed by Western blotting for HSP70 and HSP40 (Figure 2.5). Unlike arsenite, which also activates multiple other cellular stress responses, our transcriptome

analysis (see **Table 2.2**, **Figure 2.4**) shows that STA-9090 robustly induces (>2-fold) a set of only 18 genes, 13 of which are well-established HSR genes. Pretreatment with dox to induce dn-cHSF1 expression completely inhibits the induction of these 13 HSR genes, with no effects on the other five genes activated by STA-9090 treatment. The remaining five genes that remain upregulated in response to STA-9090 independent of dn-cHSF1 activation appear to be unrelated to heat shock, leading us to attribute their upregulated expression levels to other effects of STA-9090 treatment. Thus, dn-cHSF1 induction provides a mechanism to uncouple the direct consequences of HSP90 inhibition from those effects mediated by compensatory HSR activation. This method should prove valuable in the numerous cell systems where no window exists to inhibit HSP90 (or other cytosolic chaperones) without simultaneously inducing the HSR.

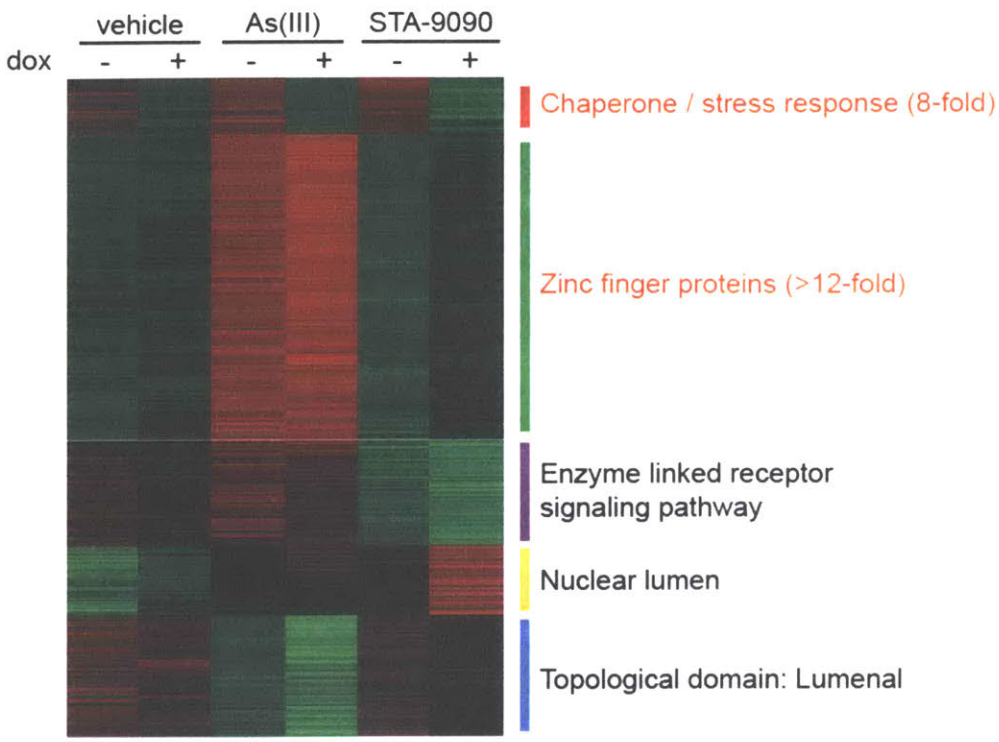

Figure 2.4 | Transcriptional profiling of the selective inhibition of HSF1 upon dn-cHSF1 activation

A five-node K-means clustergram performed on genes showing an experiment-wide ANOVA of <0.001 in array data obtained from HEK293T-REx cells inducibly expressing dn-cHSF1 following pretreatment with vehicle or dox (18 h; 1 μg/mL) and then subjected to HSR activation by treatment with arsenite (6 h; 100 μM) or STA-9090 (6 h; 100 nM). Red corresponds to high relative transcript levels, and green corresponds to low relative transcript levels. Listed next to each node is the strongest gene ontology (GO) assignment based on enrichment scores from DAVID GO analysis of gene sets. Labels in red correspond to GO assignments with an enrichment score >2 fold over expected.

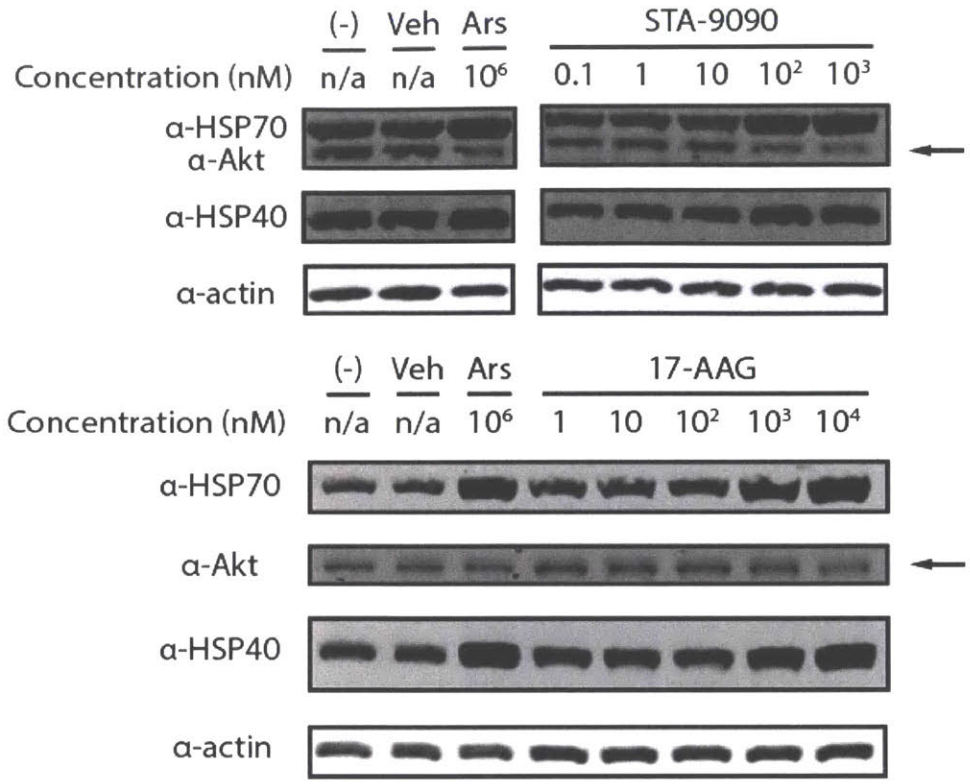

Figure 2.5 | HSP90 inhibitor-mediated activation of the heat shock response.

Immunoblot of HEK293T-Rex cells upon treatment with increasing concentrations of the HSP90 inhibitors STA-9090 and 17-AAG, showing that the heat shock response is activated even at concentrations incapable of inducing Akt degradation, a marker for successful HSP90 inhibition.

Table 2.2 | Summary of RNA array data measuring the specificity of dn-cHSF1-mediated

gene	arsenite	arsenite + dn-cHSF1	HSP90 inhibitor	HSP90 inhibitor + dn-cHSF1
		cytosolic chaperones		
HSPA1A; HSPA1B	**13.15**	0.52	**18.68**	1.19
HSPA6; HSPA7	**18.04**	0.66	**8.47**	0.62
HSPA6	**13.88**	0.85	**7.04**	0.72
DNAJB1	**3.43**	0.38	**3.31**	0.38
HSPH1	**2.28**	1.29	**2.92**	0.95
DNAJA1	**1.76**	0.94	**1.82**	0.9
HSP90AA1	**1.39**	0.66	**1.78**	0.69
		oxidative stress genes		
HMOX1	**27.67**	**16.92**	1.15	0.8
CTH	**8.47**	**8.37**	1.3	3.47
MT1X	**25.55**	**24.8**	1.16	1.33
MT1F	**16.2**	**19.01**	1.13	1.44
		zinc metalloproteins and transporters		
ZNF195	**2.09**	**2.29**	1.22	1.34
ZNF33A	**1.4**	**1.33**	1.12	1.01
ZFP69B	**5.65**	**6.93**	1.09	1.8
ZNF10	**2.19**	**2.13**	1.07	1.16
ZMYM3	**0.45**	**0.43**	1.06	0.85
DPF2	**1.51**	**1.52**	0.98	0.98
SLC30A2	**6**	**12.89**	0.96	1.3
ZNF674	**2.33**	**2.21**	0.73	1.07
		endoplasmic reticulum chaperones		
HSPA5 (GRP78)	0.81	0.83	0.77	0.96
GRP94	0.88	0.84	0.98	1.01
DNAJB11 (ERDJ3)	0.81	0.89	0.85	0.81
DNAJB9 (ERDJ4)	1.06	0.99	1	1.22
DNAJC10 (ERDJ5)	1.01	0.8	0.99	0.93
HYOU1	1.04	0.98	1.13	1.05
		disulfide redox		
PDIA3	1.15	1	1.01	1.08
PDIA4	1.04	1	1.15	1.17
PDIA6	0.85	0.75	0.9	0.99
ERO1L	0.93	0.83	0.97	1.02
ERO1LB	1.11	1.24	0.73	1.03

[a]Bolded text indicates an ANOVA p value <0.001 relative to vehicle-treated cells. All unbolded text indicates no significant change with respect to vehicle-treated cells.

2.4.3 Development of a Convenient and Broadly Applicable Small Molecule-Regulated Method to Inhibit Endogenous HSF1

With our highly potent and selective dn-cHSF1 construct in hand, we next anticipated that we could create a small molecule-regulated version of the dn-cHSF1 protein by fusion to an appropriate DD. To test this concept, we fused the Arg12Tyr/Tyr100Ile variant of *E. coli* DHFR, whose levels can be regulated in mammalian cells by the small molecule trimethoprim (TMP),[38] to the N-terminus of dn-cHSF1 (**Figure 2.6**). The addition of TMP to HEK293T-REx cells expressing this DHFR.dn-cHSF1 fusion stabilizes the entire protein (**Figure 2.7**), effectively inhibiting the HSR induced by a diverse array of stressors. For example, as shown in **Figure 2.8**, TMP addition abrogates the HSF1-mediated transcriptional upregulation of *HSP40* induced by arsenite, the HSP90 inhibitor STA-9090, heat shock, and the HSP70 inhibitor MAL3-101.[44] These results are recapitulated at the protein level by immunoblotting for HSR target proteins (**Figure 2.9**). Control experiments with DHFR.YFP-expressing cells confirm that these effects are mediated by the DHFR.dn-cHSF1 construct.

Three critical potential advantages of DD regulation of transcription factors are (1) high dosability, (2) rapid activation upon small molecule addition owing to the post-translational mechanism of regulation, and (3) ease of transportability into any model system of interest with minimal optimization.[36, 40] Indeed, we observe that TMP addition dose-dependently suppresses HSF1 activity, allowing us to titer in distinct levels of HSF1 inhibition at will (**Figure 2.10).** The timecourse of DHFR.dn-cHSF1 activation is also rapid. Within <2 h of TMP addition, we already observe the capacity to significantly prevent arsenite-induced, HSF1-mediated induction of HSR genes, including *HSP40* (**Figure 2.11**). Most importantly, our methodology is readily transported into various cell systems, including nonhuman MDCK cells (**Figure 2.12**), with no or minimal optimization. Cumulatively, these results highlight key advantages of DD regulation of dn-cHSF1, providing the first generally applicable method to inhibit endogenous HSF1 with high selectivity and potency in a small molecule-dependent manner in diverse living systems.

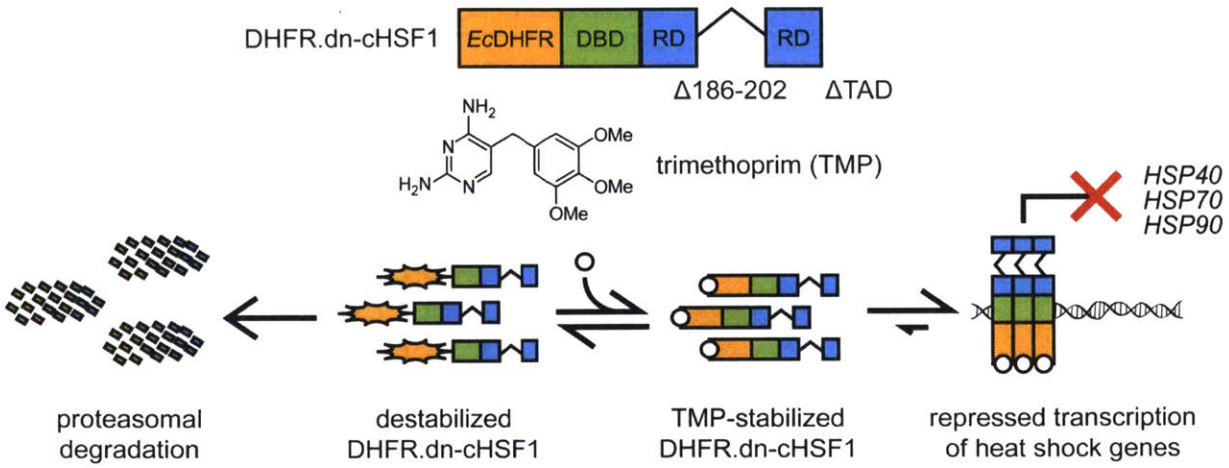

Figure 2.6 | Schematic of destabilizing domain-mediated regulation of dn-cHSF1

Model showing the trimethoprim (TMP)-dependent regulation of the DHFR.dn-cHSF1 fusion protein. The structure of TMP is shown.

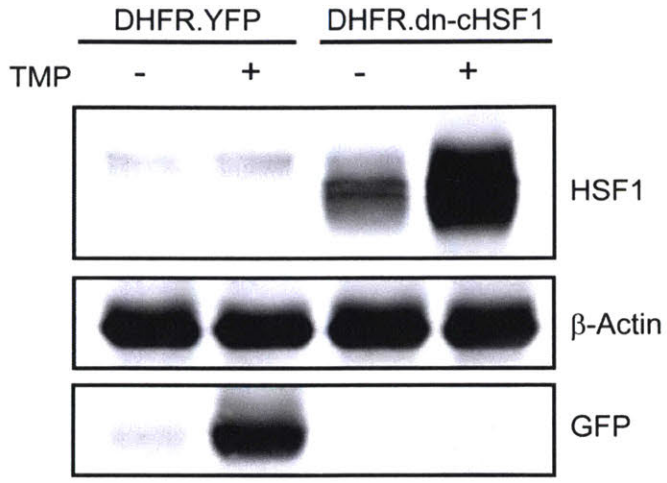

Figure 2.7 | Immunoblot of HEK293T-REx cells conditionally expressing DHFR.YFP or DHFR.dn-cHSF1

The stabilizing ligand TMP (10 μM) was added 18 h prior to harvest.

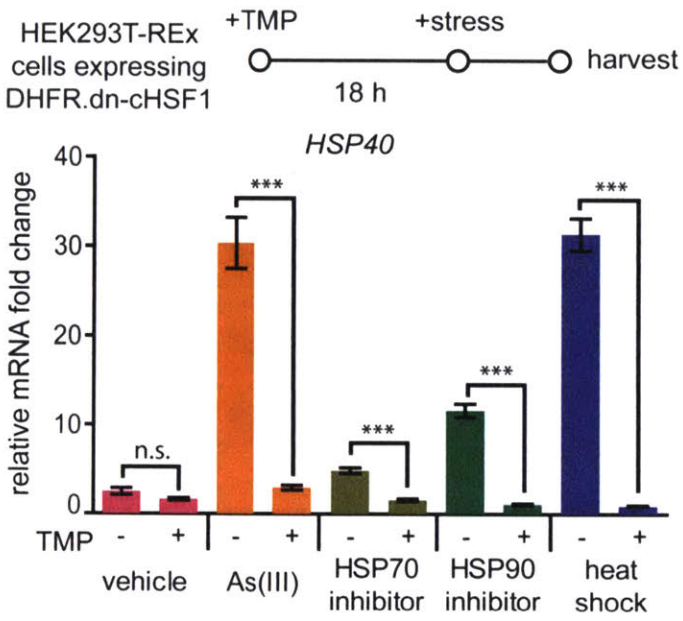

Figure 2.8 | qPCR analysis of *HSP40* in HEK293T-REx cells expressing DHFR.dn-cHSF1

Vehicle (DMSO) or TMP (10 μM) was added for 12 h, followed by treatment with arsenite (100 μM; 6 h), HSP90 inhibitor STA-9090 (100 nM; 6 h), HSP70 inhibitor MAL3-101 (10 nM; 6 h), or heat shock (42 °C; 2 h). qPCR data are presented as fold-increase relative to vehicle-treated cells expressing DHFR.YFP. Error bars represent SEM from biological replicates (n = 3). ***p < 0.0001

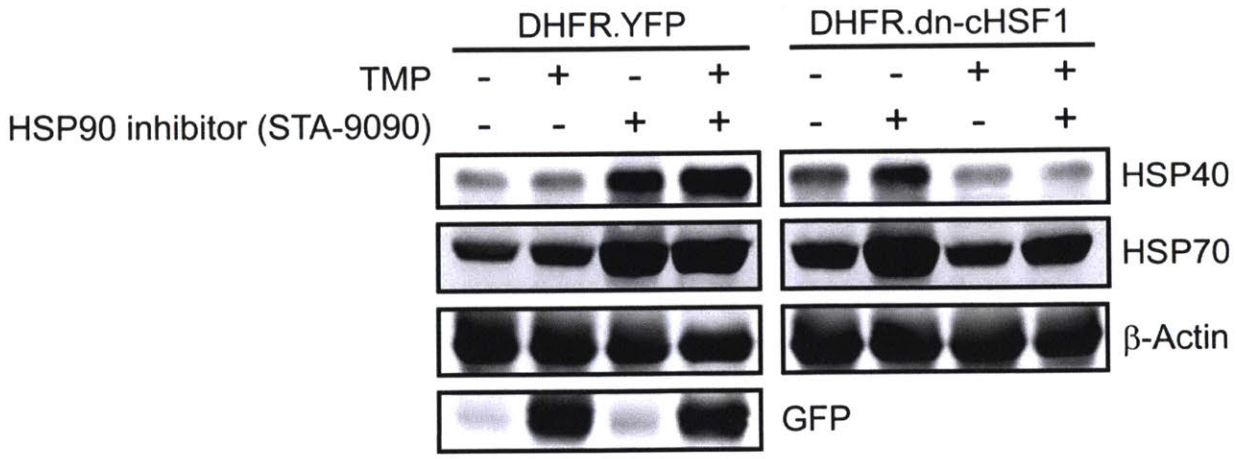

Figure 2.9 | Functional analysis of destabilized domain-regulated DHFR.dn-cHSF1 construct at protein level

Immunoblot of HEK293T-REx cells inducibly expressing DHFR.dn-cHSF1 following pretreatment with vehicle or TMP (12 h; 10 μM) and then HSR activation by treatment with HSP90 inhibitor STA-9090 (6 h; 100 nM).

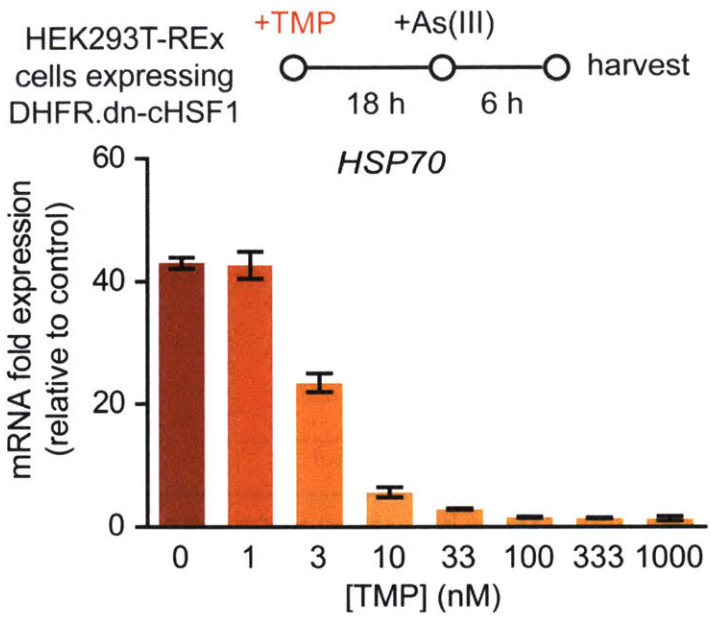

Figure 2.10 | Stabilizing ligand dose response curve for DHFR.dn-cHSF1

qPCR analysis of *HSP40* in HEK293T-REx cells expressing DHFR.dn-cHSF1 pretreated with increasing concentrations of TMP for 18 h prior to HSR activation by arsenite treatment (100 μM; 6 h). Data are reported as the mean ± 95% confidence interval relative to vehicle-treated HEK293T-REx cells.

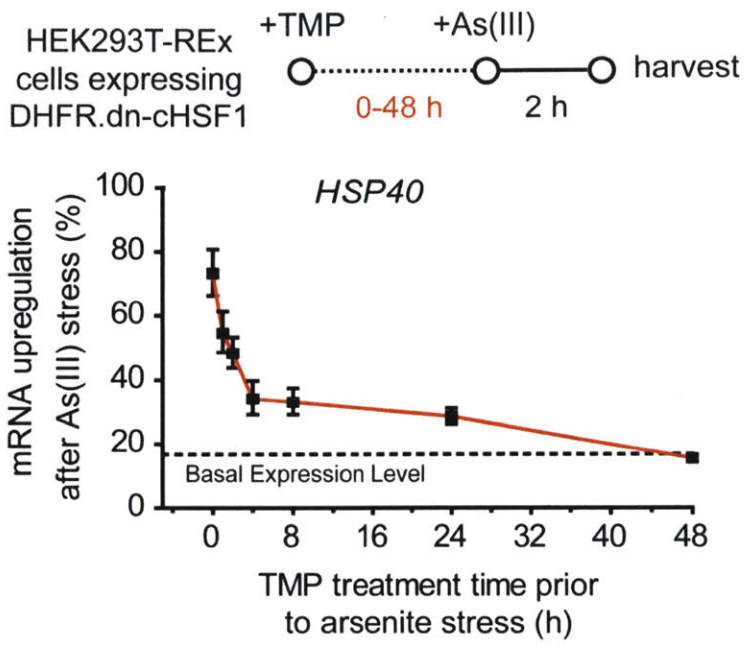

Figure 2.11 | Time course of DHFR.dn-cHSF1 inactivation of heat shock response

qPCR analysis of *HSP40* in HEK293T-REx cells expressing DHFR.dn-cHSF1 pretreated with TMP (10 μM) for the indicated times prior to HSR activation by arsenite treatment (100 μM; 2 h). The expression of *HSP40* was normalized to the maximal expression observed upon arsenite treatment with no TMP pretreatment. Data are reported as the mean ± 95% confidence interval relative to vehicle-treated HEK293T-REx cells.

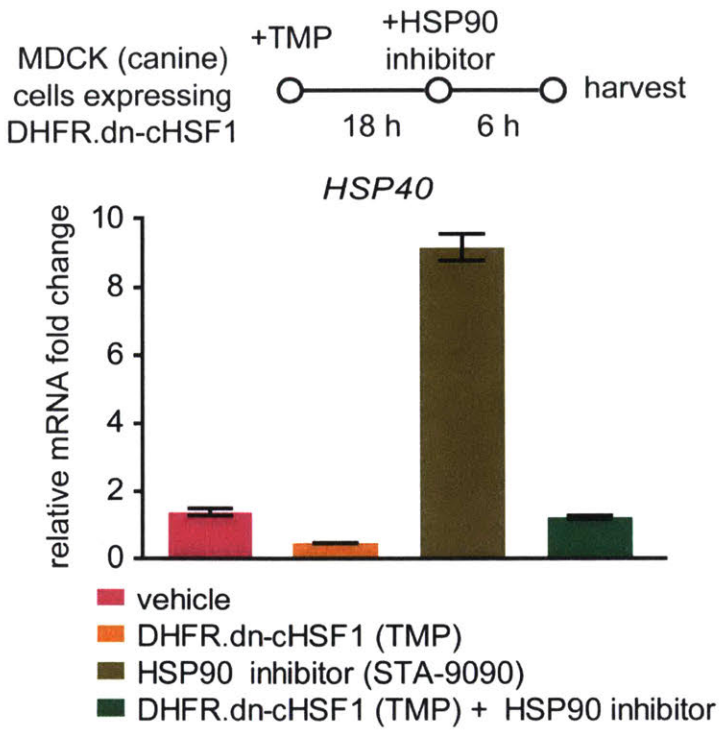

Figure 2.12 | Functional test of DHFR.dn-cHSF1 construct in non-human cells

qPCR analysis of HSF1 target genes in MDCK cells expressing DHFR.dn-cHSF1, pretreated with TMP (10 μM; 18 h) and then treated with arsenite (100 μM; 6 h) to activate the HSR. Data are presented as the mean ± 95% confidence interval relative to vehicle-treated cells.

2.4.4 Orthogonal, Small Molecule-Mediated Up- and Down-Regulation of the HSR in the Absence of Stress

Acute activation of dn-cHSF1 by TMP treatment prevents HSF1 activation in response to stress. Interestingly, and consistent with previous work,[23] we find that chronic, long-term activation of DHFR.dn-cHSF1 by TMP treatment constitutively depletes cellular chaperones by inhibiting basal HSF1 activity (**Figure 2.13**). The capacity to constitutively downregulate the cytosolic proteostasis network by DHFR.dn-cHSF1 activation, if integrated with an orthogonal method for stress-independent HSF1 activation, would allow us to create individual cell lines in which the cytosolic proteostasis network can be inducibly up- or down-regulated by treatment with orthogonal small molecules (**Figure 2.14**). Such cell lines would be valuable for testing effects of HSF1 modulation on the proteostasis of particular normal and disease-related protein variants.

Methodology for the small molecule-mediated activation of cHSF1 using the FKBP DD, which is stabilized by the small molecule Shield-1, was recently reported.[36, 41] Because the DHFR and FKBP DDs are regulated by distinct small molecules, we predicted that expressing both DHFR.dn-cHSF1 and FKBP.cHSF1 in a single cell line would allow us to inhibit or activate HSF1 in a stress-independent manner in a single population of cells. Upon generating HEK293T-REx cells expressing both constructs, termed HEK293$^{DD.HSR}$ cells for DD-regulated control of the HSR, we observe that Shield-1 treatment robustly activates HSF1, as illustrated by the upregulation of *HSP90*, *HSP70*, and *HSP40* (**Figure 2.15**). In contrast, TMP treatment activates DHFR.dn-cHSF1 to chronically deplete cytosolic chaperones or inhibit stress-mediated HSF1 activation. Because both FKBP.cHSF1 and DHFR.dn-cHSF1 are readily transported into any cell line of interest with minimal optimization,[36] this methodology provides a convenient mechanism to directly test the potentially contrasting effects of small molecule-mediated HSF1 activation versus small molecule-mediated HSF1 inhibition in disease model systems ranging

from the protein misfolding disorders to cancer. Similar results are observed upon dox or Shield-1 treatment in cells expressing tetracycline repressor-regulated dn-cHSF1 and FKBP.cHSF1 (which we term HEK293^{HSR} cells; **Figure 2.16**).

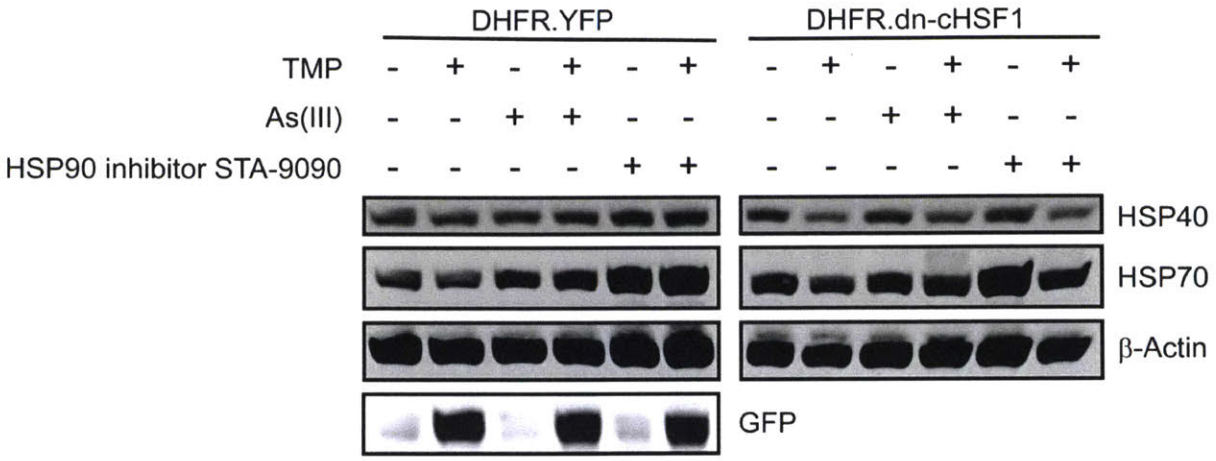

Figure 2.13 | Chronic expression of dn-cHSF1 reduces chaperone protein levels

Immunoblot showing the depletion of cellular chaperones upon chronic activation of DHFR.dn-cHSF1 by treatment with TMP (10 μM; 72 h).

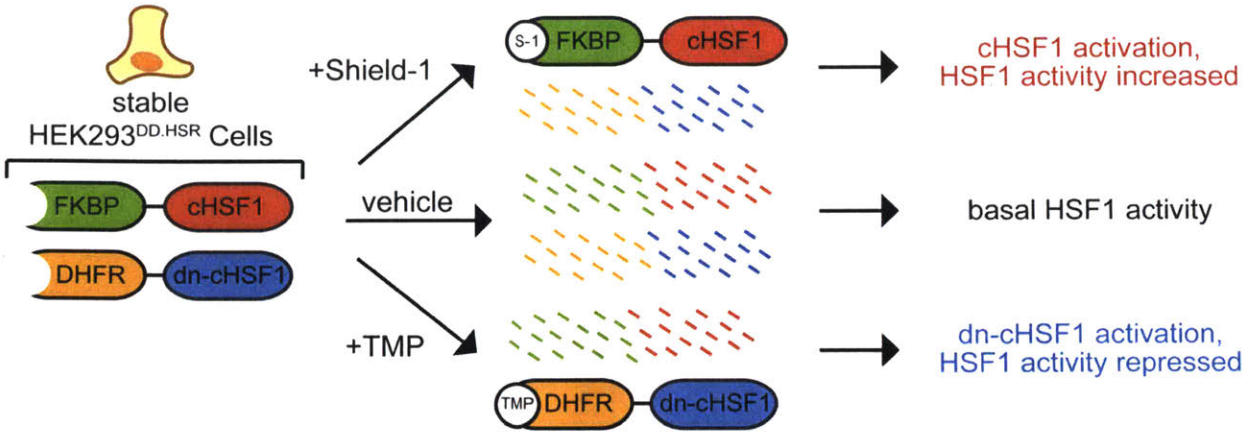

Figure 2.14 | Stress-independent upregulation and depletion of chaperones can be combined in a single cell system

Illustration showing the incorporation of both DHFR.dn-cHSF1 and FKBP.cHSF1 into a single cell line (HEK293$^{DD.HSR}$ cells) and expected outcomes for proteostasis environment in the cytosol.

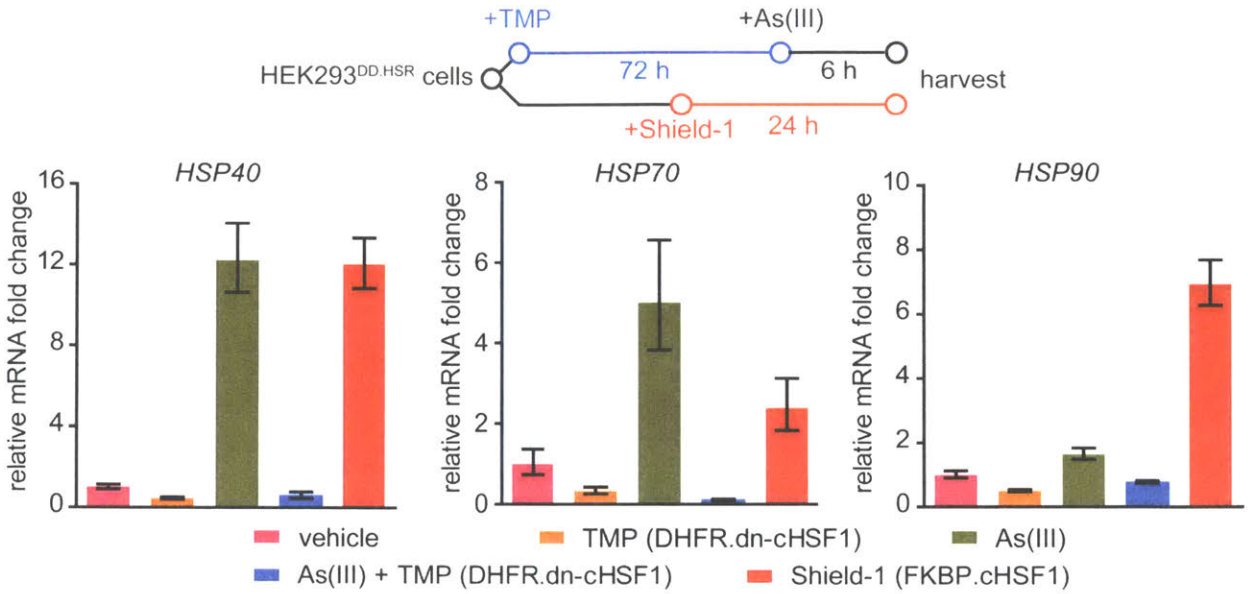

Figure 2.15 | qPCR analysis of HSF1 target genes in HEK293$^{DD.HSR}$ cells expressing both TMP-regulated DHFR.dn-cHSF1 and Shield-1-regulated FKBP.cHSF1

c cells were treated with TMP (10 μM; 78 h), Shield-1 (1 μM; 24 h), arsenite (100 μM; 6 h), or pretreated with TMP (10 μM; 72 h) followed by arsenite (100 μM; 6 h). Data are reported as the mean ± 95% confidence interval relative to vehicle-treated HEK293T-REx cells.

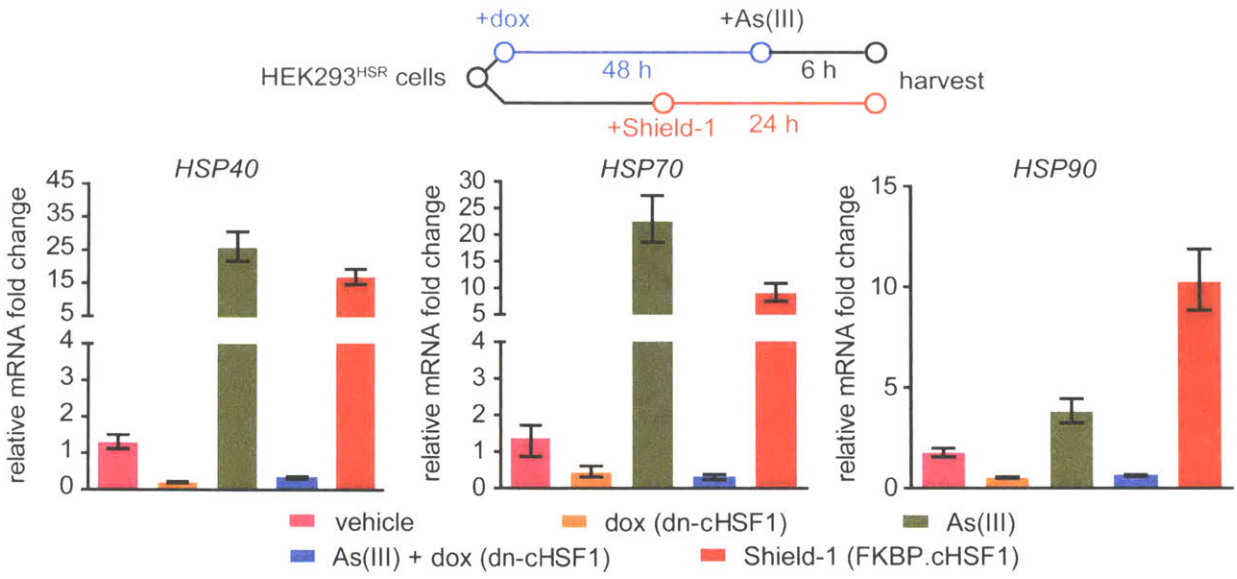

Figure 2.16 | qPCR analysis of HSF1 target genes in HEK293^{HSR} cells expressing both dox-regulated dn-cHSF1 and Shield-1-regulated FKBP.cHSF1.

HEK293^{HSR} cells were treated with dox (1 µg/mL; 54 h), Shield-1 (1 µM; 24 h), arsenite (100 µM; 6 h), or pretreated with dox (1 µg/mL; 48 h) followed by arsenite (100 µM; 6 h). Data are reported as the mean ± 95% confidence interval relative to vehicle-treated HEK293T-REx cells.

2.4.5 Effects of HSF1 Modulation on the Behavior of Cytosolic Chaperone Clients

We next evaluated how small molecule-mediated remodeling of the cytosolic proteostasis network in HEK293^{HSR} cells influences established model cytosolic chaperone client proteins, such as firefly luciferase (FLuc).[45, 46] We expected that upregulating cytosolic proteostasis mechanisms via FKBP.cHSF1 activation might enhance FLuc activity, while chronic dn-cHSF1 activation to deplete the cytosolic proteostasis network would reduce FLuc activity. Surprisingly, we observe that inhibiting endogenous HSF1 and lowering chaperone levels by inducing dn-cHSF1 significantly increases FLuc activity >2.5-fold, while Shield-1 treatment to activate HSF1 and upregulate the cytosolic proteostasis network marginally reduces FLuc activity (**Figure 2.17**). Interestingly, accounting for steady-state levels of cytosolic FLuc (**Figure 2.18**) reveals that the specific activity of FLuc in basal, activated or inhibited HSF1 cellular environments is not significantly impacted by HSF1 activation or inhibition. The enhanced FLuc activity upon HSF1 inhibition cannot be attributed to globally altered protein translation, because the extent of metabolic labeling of newly synthesized proteins remains unchanged by the presence or absence of dn-cHSF1 (**Figure 2.19**). Thus, the observed changes in FLuc activity may be related to modified quality control induced through chronic inhibition of endogenous HSF1, though other mechanisms could certainly also play a role. These results highlight the complex and potentially unpredictable effects of an altered proteostasis network for some chaperone client proteins. The trends we observe for FLuc activity are consistent with recent work on the maladapted stress response by Balch and co-workers, who identified HSF1 knockdown as a potential mechanism to enhance proteostasis of misfolding globular and membrane proteins associated with genetic loss-of-function disorders.[14]

We next profiled the effects of HSF1 inhibition versus activation on the intrinsically disordered, aggregating protein polyQ$_{67}$-tdTomato,[41, 45] a protein consisting of 67 glutamine residues fused to fluorescent tdTomato for visualization (**Figure 2.20**). Polyglutamine proteins

like this one are a classic example of the intrinsically disordered proteins associated with the pathology of numerous human diseases, including Huntington's disease. Various groups have speculated that modulating the proteostasis network could be a viable therapeutic strategy for such diseases.[13] Immunoblotting for polyQ$_{67}$-tdTomato from both the soluble and insoluble fractions of cell lysates suggests that the protein aggregates more aggressively in HSF1-inhibited environments (**Figure 2.21**, note also the presence of a cleaved tdTomato band that has been previously observed for polyQ-fluorescent protein fusions and that is unaltered by modulating HSF1 activity[47]). Indeed, we find that under conditions of HSF1 inhibition the aggregation of polyQ$_{67}$-tdTomato is significantly enhanced, as determined by the appearance of fluorescent puncta identified by PULSA flow cytometry (**Figure 2.22**, **Figure 2.23**, and **Table 2.3**).[48] HSF1 activation, on the other hand, does not significantly affect puncta formation, resulting in a similar amount of disperse fluorescence to untreated samples. Thus, while we find that HSF1 inhibition and consequent chaperone depletion can possibly be beneficial for soluble but poorly folding globular proteins like FLuc, these results suggest that inhibiting HSF1 and repressing the cytosolic proteostasis network reduces the cell's capacity to handle aggregation-prone proteins.

Cumulatively, these data highlight the potentially contrasting consequences of HSF1 inhibition for different classes of misfolding proteins. For a soluble chaperone client, we observe an unexpected increase in enzyme activity upon HSF1 inhibition, an effect that could prove useful in certain protein misfolding-related diseases.[14] In contrast, HSF1 inhibition significantly enhances deleterious aggregate formation for an intrinsically disordered protein, and approaches that upregulate proteostasis network activity may prove more useful. These findings highlight the importance of considering the therapeutic relevance of both possibilities, depending on the protein misfolding-related disorder of interest, and lend credence to the concept that small molecule-mediated inhibition of HSF1 is a strategy that should be tested in

cytosolic protein misfolding-related disorders.[14] Such testing should benefit greatly from our introduction of DHFR.dn-cHSF1.

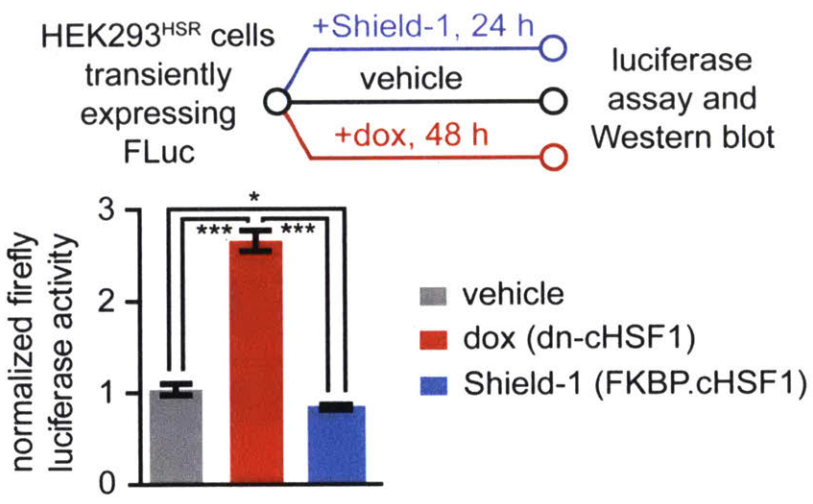

Figure 2.17 | Activity of a chaperone client is higher in a chaperone-depleted environment

Experimental workflow showing FLuc activity measured in transiently transfected HEK293T-REx^{HSR} cells after treatment with vehicle, dox to induce HSF1 inhibition (1 µg/mL; 48 h), or Shield-1 to activate cHSF1 (1 µM; 48 h). Error bars represent SEM from biological replicates (n = 3). Significance was calculated using the paired Student's two-tailed *T*-test. *$p < 0.05$, ***$p < 0.005$.

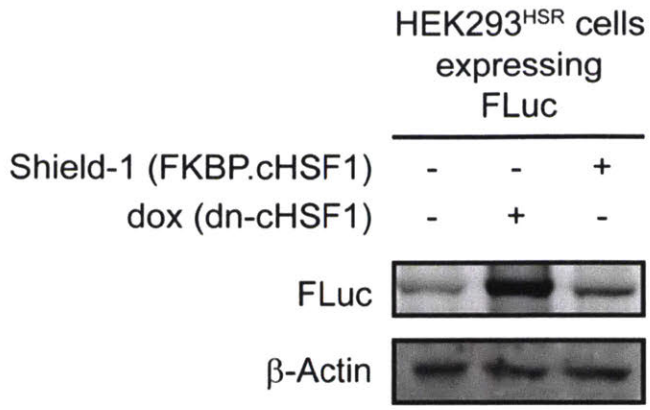

Figure 2.18 | Protein levels of a chaperone client are elevated in a chaperone-depleted environment

Immunoblot to assay steady-state levels of FLuc after transiently transfected HEK293T-REx^{HSR} cells after treatment with vehicle, dox to induce HSF1 inhibition (1 μg/mL; 48 h), or Shield-1 to activate cHSF1 (1 μM; 48 h).

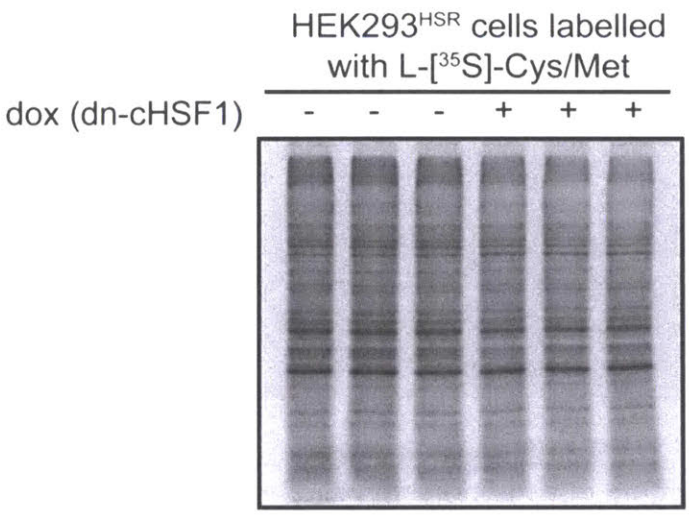

Figure 2.19 | Global protein synthesis remains constant after chronic expression of dn-cHSF1

Auto-radiogram of [^{35}S]-labelled protein from HEK293^{HSR} lysates expressing dn-cHSF1 (dox; 1 µg/mL, 48 h) or no treatment prior to 15 min metabolic labelling with L-[^{35}S]-methionine.

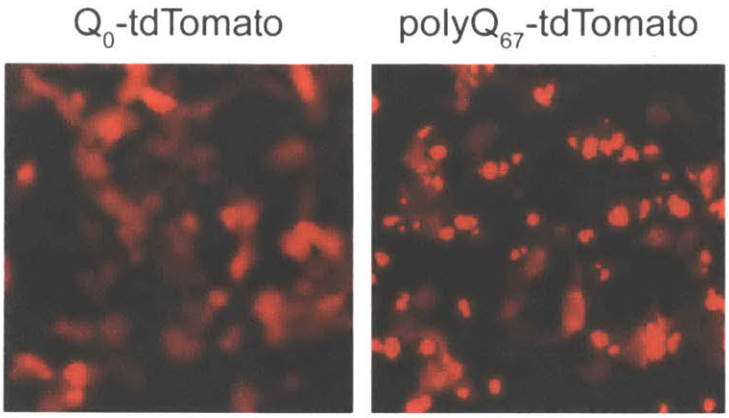

Figure 2.20 | polyQ-tagging of a fluorescent protein causes aggregation, resulting in fluorescent puncta when expressed in cells

Epifluorescent microscopy images of HEK293^{HSR} cells transiently transfected with plasmids encoding Q_0-tdTomato (left) or $polyQ_{67}$-tdTomato (right))

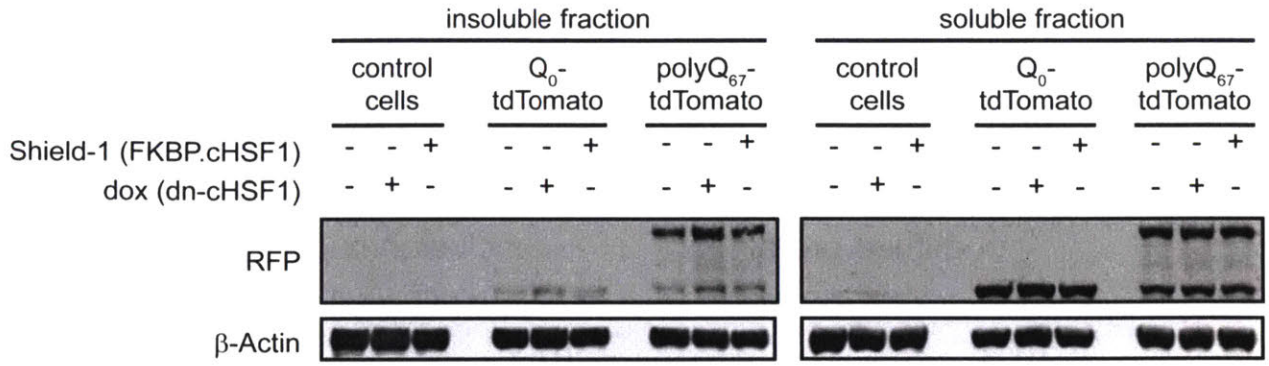

Figure 2.21 | polyQ-protein aggregation is higher after chronic expression of dn-cHSF1

Immunoblot of soluble and insoluble fractions from HEK293^{HSR} cells transfected with Q_0-tdTomato or $polyQ_{67}$-tdTomato, and then treated with vehicle, dox to induce HSF1 inhibition (1 µg/mL; 48 h), or Shield-1 to activate cHSF1 (1 µM; 48 h).

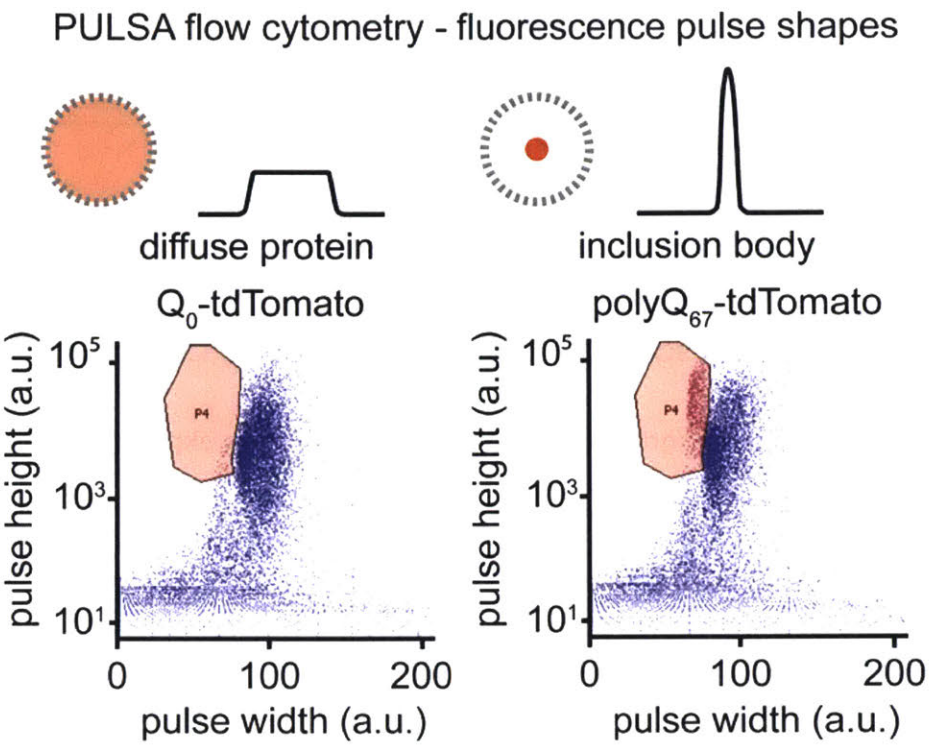

Figure 2.22 | Schematic and representative data for PULSA flow cytometry method to detect fluorescent inclusion bodies in living cells

Diagram of PULSA flow cytometry, which utilizes fluorescence pulse height and pulse width information to identify inclusion bodies in live cells. Shown are representative plots of HEK293^{HSR} cell populations transiently transfected with Q_0-tdTomato (left) and polyQ$_{67}$-tdTomato (right), with gating for inclusion body population defined as the region bounded by the orange line. Purple dots in this region are defined as live cells with inclusion bodies, while all other events on the plot are defined as live cells.

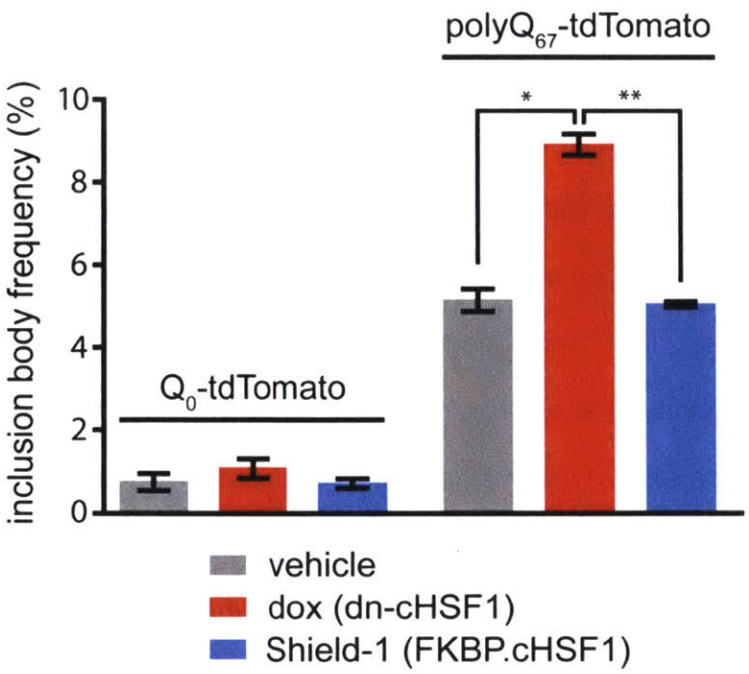

Figure 2.23 | Chaperone depletion exacerbates polyQ-mCherry aggregation

Quantification of inclusion body frequency measured by PULSA flow cytometry analysis of HEK293^{HSR} cells transfected with Q_0-tdTomato or polyQ$_{67}$-tdTomato, and then treated with vehicle, dox to induce HSF1 inhibition (1 µg/mL; 48 h), or Shield-1 to activate cHSF1 (1 µM; 48 h). Error bars represent SEM from 10,000 recorded live cell events performed on biological replicates (n = 3). *p < 0.05, **p < 0.01.

2.4.6 Concluding Remarks

HSF1 is a high-profile target for both diseases of proteostasis and cancer. In the absence of highly selective and potent small molecule-regulated methods to inhibit HSF1 in disease model systems, understanding the beneficial and/or deleterious consequences of HSF1 inhibition remains challenging. Here, we report the development of a potent and exquisitely selective small molecule-regulated inhibitor of endogenous HSF1. Acute activation of our inhibitor abrogates the HSF1-mediated, stress-induced HSR, while chronic activation constitutively depletes cytosolic chaperones. This small molecule-regulated methodology is valuable for mechanistic experiments, where it permits the uncoupling of the direct effects of environmental toxins and chaperone inhibitors from the indirect effects of the compensatory HSR induced by these compounds. Furthermore, integration of our method with a previously reported small molecule-based technique to activate HSF1[36] permits the generation of enhanced or repressed cytosolic proteostasis networks using orthogonal, stress-independent small molecules in a single population of cells. Intriguingly, we find that the consequences of HSF1 inhibition versus activation on proteostasis are not always intuitive. For a model chaperone-dependent soluble protein, HSF1 inhibition has the beneficial effect of significantly increasing enzyme activity, while HSF1 activation very modestly reduces enzyme activity. In contrast, for the alternative case of an intrinsically disordered, aggregation-prone protein, HSF1 inhibition enhances deleterious protein aggregation. These results highlight the value of selective and potent chemical biology methods to inhibit HSF1, allowing hypothesis testing in disease-relevant systems to guide rational therapeutic development efforts and elucidation of the critical roles of HSF1 in tumorigenesis and cancer progression.

2.5 Methods

2.5.1 Reagents, Plasmids, and Antibodies

Sodium arsenite 0.1 N standardized solution was purchased from Alfa Aesar, STA-9090 was purchased from MedChem Express, MAL3-101 was purchased from ChemTech, Inc., and Shield-1 was purchased from ClonTech. FLuc-GFP.pCIneo was a generous gift from Prof. F.-U. Hartl (Max Planck Institute).[45] p-N1-Q_0-tdTomato and p-N1-polyQ_{67}-tdTomato plasmids were a generous gift from Prof. R. Morimoto (Northwestern).[41] To construct dn-cHSF1.pENTR1A and DHFR.dn-cHSF1.pENTR1A, the DNA fragment corresponding to dn-cHSF1 was PCR-amplified using DNA oligonucleotides aaaaaaggtaccaccatggatctgcccgtggg and aaaaaagcggccgcctacaggcaggctacgctga as primers and cHSF1.pENTR1A as the template.[36] The PCR product was digested with *Kpn*I and *Not*I, and then cloned into *Kpn*I- and *Not*I-digested pENTR1A (Life Technologies) or DHFR.YFP.pENTR1A vectors.[38] Genes of interest in pENTR1A were shuttled into appropriate destination vectors using LR clonase II-mediated recombination (Life Technologies). The following antibodies were used: mouse monoclonal anti-β-actin and rabbit polyclonal anti-HSF1 from Sigma, anti-HSP70/72 and anti-HSP40/Hdj1 from Enzo Life Sciences, rabbit monoclonal anti-HSP90 from Cell Signaling, rabbit polyclonal anti-RFP from Evrogen, and rabbit polyclonal anti-GFP from GeneTex.

2.5.2 Cell Culture

HEK293T-REx (Life Technologies) and MDCK cells were cultured at 37 °C in a 5% CO_2atmosphere in DMEM (CellGro) supplemented with 10% fetal bovine serum (CellGro) and 1% penicillin/streptomycin/glutamine (CellGro). Vesicular stomatitits virus glycoprotein (VSV-G) pseudotyped lentiviral particles encoding dox-inducible dn-cHSF1 and DHFR.dn-cHSF1 were produced by co-transfecting 293FT cells with the structural plasmids necessary for virus production (Rev, RRE, and VSVG) along with the appropriate pLenti-DEST lentivirus constructs. Cells were transfected using Lipofectamine 2000 (Life Technologies) for 24 h, after which the

media was removed and replaced with fresh media. Media containing viral particles was collected at 48 h and again at 72 h post-transfection and cell debris was removed by centrifugation at 500 × g for 10 min. Viral supernatant was then aliquoted and stored at –80 °C until use. During lentivirus transductions, cells were treated with 4 μg/mL polybrene (Sigma). Lentiviral particles were transduced into either MDCK cells, HEK293T-REx cells, or previously described HEK293T-REx cells already stably expressing FKBP.cHSF1.[36] Stable cell lines were selected by culturing in complete medium containing G418, blasticidin, Zeocin, hygromycin, and/or puromycin, as appropriate, prior to single-colony selection and characterization. Transient transfections of polyglutamine and FLuc-GFP.pCIneo constructs were performed using polyethylenimine.

2.5.3 Immunoblotting

Proteins were separated by SDS-PAGE and then transferred to nitrocellulose membranes. Following blocking and incubation with appropriate primary antibodies, membranes were incubated with 680 or 800 nm fluorophore-labeled secondary antibodies (Li-COR Biosciences) prior to detection using a LI-COR Biosciences Odyssey Imager. Band intensity quantification was performed in Image Studio Lite (LI-COR Biosciences).

2.5.4 Quantitative RT-PCR

The relative mRNA expression levels of selected heat shock response genes were measured using quantitative RT-PCR. Cells were treated as described at 37 °C, harvested by scraping, washed with phosphate-buffered saline (CellGro), and then RNA was extracted using the EZNA Total RNA Kit I (Omega). qPCR reactions were performed on cDNA prepared from 1000 ng of total cellular RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The Fast Start Universal SYBR Green Master Mix (Roche) and appropriate primers purchased from Integrated DNA Technologies (**Table 2.1**) were used for amplifications (6 min at 95 °C then 45 cycles of 10 s at 95 °C, 30 s at 60 °C) in a Light Cycler 480 II Real-Time

PCR machine. Primer integrity was assessed by a thermal melt to confirm homogeneity and the absence of primer dimers. Transcripts were normalized to the housekeeping genes *Rplp2* and all measurements were performed at least in triplicate. Data were analyzed using the LightCycler® 480 Software, Version 1.5 (Roche) and data are reported as the mean ±95% confidence interval.

2.5.5 Whole Genome Microarrays

HEK293T-REx cells expressing dox-inducible dn-cHSF1 were treated for 18 h with vehicle or 1 µg/mL dox prior to HSF1 activation by treatment with arsenite (100 µM for 4 h) or STA-9090 (10 nM for 8 h), all in biological triplicate. Cells were harvested and RNA was extracted using the RNeasy Plus Mini Kit (Qiagen). RNA quality was confirmed using an Advanced Analytical Fragment Analyzer, after which 250 ng of total RNA was prepared for microarray analysis using the Nugen Applause 3' Amp Kit (NuGen #5100). 4 µg of labeled cDNA was hybridized overnight (17 h) to Human Primeview Arrays (Affymetrix #901837) following standard Affymetrix protocols. Microarrays were scanned using an Affymetrix GeneChip Scanner 3000 7G Series instrument. Data were extracted using the Affymetrix Expression Console and analyzed using the Transcriptome Analysis Console v3.0. K-Means clustering was performed on genes showing an experiment-wide ANOVA of <0.001 using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). Data were gene normalized before clustering using 5 nodes and a Euclidean distance similarity metric. Data were visualized using Java TreeView (http://jtreeview.sourceforge.net).1 Gene ontology analyses were performed using DAVID (http://david.abcc.ncifcrf.gov/summary.jsp) on genes associated to each node.

2.5.6 Luciferase Activity Assays

After pretreatment with the appropriate small molecule and/or stressor, cells were lysed by adding luciferase lysis buffer (Promega) directly to the wells and incubated in the dark at room temperature for 15 min. Following a 5 min spin to remove insoluble cell debris, 50 µL of lysate was added to a white, opaque 96-well plate with 50 µL of the Bright-Glo Luciferase Assay System buffer (Promega). Following a 15 s double-orbital mix cycle, luminescence was recorded on a BioTek Synergy H1 Hybrid Reader. Measurement times using luminescence fiber were set to 0.4 s integration time per well. A four-point mean maximum calculation was used to determine the luminescence. To determine specific activities, the luminescence values (FLuc activities) were divided by FLuc band intensity quantified from immunoblots by densitometry using the LI-COR Image Studio Lite software.

2.5.7 [^{35}S] Metabolic Labeling Experiments

HEK293^{HSR} cells transiently transfected with FLuc were seeded on poly-d-lysine-coated plates and treated with vehicle or dox at 1 µg/mL to activate dn-cHSF1 for 48 h. Cells were then starved for 30 min in DMEM + 10% FBS lacking Cys and Met. Cells were metabolically labeled in pulse medium containing [^{35}S]-Cys/Met (MP Biomedical, ~ 0.1 mCi/mL final concentration) for 15 min prior to lysis. Cells were lysed in a 1% Triton X-100 buffer. Lysates were boiled in 1X-Laemmli buffer and separated by SDS-PAGE. The gels were then dried, exposed to phosphorimager plates (GE Healthcare), and imaged with a Typhoon imager.

2.5.8 Flow Cytometry

Cells were analyzed at a slow flow rate in an LSRFortessa flow cytometer (BD Biosciences). A total of 10 000 events were collected per measurement, using a forward scatter threshold of 5000 and obtaining pulse height, area, and width parameters for each channel. For tdTomato, data were collected with the 561 nm laser and 610/20 nm bandpass filter. Flow cytometry data were analyzed using FACSDiva software (BD Biosciences).

2.6 References

1. Åkerfelt, M., Morimoto, R.I. & Sistonen, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell. Biol.* **11**, 545-555 (2010).
2. Anckar, J. & Sistonen, L. Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu. Rev. Biochem.* **80**, 1089-1115 (2011).
3. Trinklein, N.D., Murray, J.I., Hartman, S.J., Botstein, D. & Myers, R.M. The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. *Mol. Biol. Cell* **15**, 1254-1261 (2004).
4. Zuo, J.R., Baler, R., Dahl, G. & Voellmy, R. Activation of the DNA-binding ability of human heat-shock transcription factor-1 may involve the transition from an intramolecular to an intermolecular triple-stranded coiled-coil structure. *Mol. Cell. Biol.* **14**, 7557-7568 (1994).
5. Zou, J.Y., Guo, Y.L., Guettouche, T., Smith, D.F. & Voellmy, R. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell* **94**, 471-480 (1998).
6. Voellmy, R. On mechanisms that control heat shock transcription factor activity in metazoan cells. *Cell Stress Chaper.* **9**, 122-133 (2004).
7. Whitesell, L. & Lindquist, S. Inhibiting the transcription factor HSF1 as an anticancer strategy. *Exp. Op. Ther. Targets* **13**, 469-478 (2009).
8. Dai, C., Whitesell, L., Rogers, A.B. & Lindquist, S. Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. *Cell* **130**, 1005-1018 (2007).
9. Santagata, S. et al. High levels of nuclear heat-shock factor 1 (HSF1) are associated with poor prognosis in breast cancer. *Proc. Natl. Acad. Sci. USA* **108**, 18378-18383 (2011).
10. Scherz-Shouval, R. et al. The reprogramming of tumor stroma by HSF1 is a potent enabler of malignancy. *Cell* **158**, 564-578 (2014).
11. Mendillo, M.L. et al. HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell* **150**, 549-562 (2012).
12. Chen, Y.Y. et al. Targeting HSF1 sensitizes cancer cells to HSP90 inhibition. *Oncotarget* **4**, 816-829 (2013).
13. Balch, W.E., Morimoto, R.I., Dillin, A. & Kelly, J.W. Adapting proteostasis for disease intervention. *Science* **319**, 916-919 (2008).
14. Roth, D.M. et al. Modulation of the maladaptive stress response to manage diseases of protein folding. *PLoS Biol.* **12**, e10011998 (2014).
15. Au, Q.Y., Zhang, Y.J., Barber, J.R., Ng, S.C. & Zhang, B. Identification of inhibitors of HSF1 functional activity by high-content target-based screening. *J. Biomol. Screen.* **14**, 1165-1175 (2009).
16. Kim, J.A., Kim, Y., Kwon, B.M. & Han, D.C. The natural compound Cantharidin induces cancer cell death through inhibition of heat shock protein 70 (HSP70) and Bcl-2-associated athanogene domain 3 (BAG3) expression by blocking heat shock factor 1 (HSF1) binding to promoters. *J. Biol. Chem.* **288**, 28713-28726 (2013).

17. Westerheide, S.D., Kawahara, T.L.A., Orton, K. & Morimoto, R.I. Triptolide, an inhibitor of the human heat shock response that enhances stress-induced cell death. *J. Biol. Chem.* **281**, 9616-9622 (2006).
18. Santagata, S. et al. Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. *Science* **341**, 1238303 (2013).
19. Schilling, D., Kuhnel, A., Tetzlaff, F., Konrad, S. & Multhoff, G. NZ28-induced inhibition of HSF1, SP1 and NF-kappaB triggers the loss of the natural killer cell-activating ligands MICA/B on human tumor cells. *Cancer Immunol. Immunother.* (2015).
20. Smurnyy, Y. et al. DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. *Nature Chem. Biol.* **10**, 623-625 (2014).
21. Xia, W.L. et al. Modulation of tolerance by mutant heat shock transcription factors. *Cell Stress Chaper.* **4**, 8-18 (1999).
22. Voellmy, R. Dominant-positive and dominant-negative heat shock factors. *Methods* **35**, 199-207 (2005).
23. Heldens, L. et al. Co-chaperones are limiting in a depleted chaperone network. *Cell. Mol. Life Sci.* **67**, 4035-4048 (2010).
24. Heldens, L. et al. Protein refolding in peroxisomes is dependent upon an HSF1-regulated function. *Cell Stress Chaper.* **17**, 603-613 (2012).
25. Wang, J.H. et al. Blocking HSF1 by dominant-negative mutant to sensitize tumor cells to hyperthermia. *Biochem. Biophys. Res. Commun.* **290**, 1454-1461 (2002).
26. Wang, Y., Theriault, J.R., He, H., Gong, J. & Calderwood, S.K. Expression of a dominant negative heat shock factor-1 construct inhibits aneuploidy in prostate carcinoma cells. *J. Biol. Chem.* **279**, 32651-32659 (2004).
27. Verma, P., Pfister, J.A., Mallick, S. & D'Mello, S.R. HSF1 protects neurons through a novel trimerization- and HSP-independent mechanism. *J. Neurosci.* **34**, 1599-1612 (2014).
28. Evans, C.G., Chang, L. & Gestwicki, J.E. Heat shock protein 70 (Hsp70) as an emerging drug target. *J. Med. Chem.* **53**, 4585-4602 (2010).
29. Whitesell, L. & Lindquist, S.L. HSP90 and the chaperoning of cancer. *Nature Rev. Canc.* **5**, 761-772 (2005).
30. Baldo, B. et al. A screen for enhancers of clearance identifies Huntingtin as a heat shock protein 90 (Hsp90) client protein. *J. Biol. Chem.* **287**, 1406-1414 (2012).
31. Wang, Y. & McAlpine, S.R. Combining an Hsp70 inhibitor with either an N- or C-terminal Hsp90 inhibitor produces mechanistically distinct phenotypes. *Org. Biomol. Chem.* **13**, 3691-3698 (2015).
32. Burlison, J.A., Neckers, L., Smith, A.B., Maxwell, A. & Blagg, B.S.J. Novobiocin: Redesigning a DNA gyrase inhibitor for selective inhibition of Hsp90. *J. Am. Chem. Soc.* **128**, 15529-15536 (2006).
33. Koay, Y.C. et al. Chemically accessible HSP90 inhibitor that does not induce a heat shock response. *ACS Med. Chem. Lett.* **5**, 771-776 (2014).
34. Garg, G., Zhao, H.P. & Blagg, B.S.J. Design, synthesis, and biological evaluation of ring-constrained Novobiocin analogues as Hsp90 C-terminal inhibitors. *ACS Med. Chem. Lett.* **6**, 204-209 (2015).

35. Del Razo, L.M. et al. Stress proteins induced by arsenic. *Toxicol. Appl. Pharm.* **177**, 132-148 (2001).
36. Shoulders, M.D., Ryno, L.M., Cooley, C.B., Kelly, J.W. & Wiseman, R.L. Broadly applicable methodology for the rapid and dosable small molecule-mediated regulation of transcription factors in human cells. *J. Am. Chem. Soc.* **135**, 8129-8132 (2013).
37. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G.L. & Wandless, T.J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004 (2006).
38. Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D. & Wandless, T.J. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **17**, 981-988 (2010).
39. Miyazaki, Y., Imoto, H., Chen, L.-C. & Wandless, T.J. Destabilizing domains derived from the human estrogen receptor. *J. Am. Chem. Soc.* **134**, 3942-3945 (2012).
40. Shoulders, M.D. et al. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **3**, 1279-1292 (2013).
41. Ryno, L.M. et al. Characterizing the altered cellular proteome induced by the stress-independent activation of heat shock factor I. *ACS Chem. Biol.*, 1273-1283 (2014).
42. Ying, W.W. et al. Ganetespib, a unique triazolone-containing HSP90 inhibitor, exhibits potent antitumor activity and a superior safety profile for cancer therapy. *Mol. Cancer Ther.* **11**, 475-484 (2012).
43. Sato, S., Fujita, N. & Tsuruo, T. Modulation of Akt kinase activity by binding to Hsp90. *Proc. Natl. Acad. Sci. USA* **97**, 10832-10837 (2000).
44. Fewell, S.W. et al. Small molecule modulators of endogenous and co-chaperone-stimulated Hsp70 ATPase activity. *J. Biol. Chem.* **279**, 51131-51140 (2004).
45. Gupta, R. et al. Firefly luciferase mutants as sensors of proteome stress. *Nat. Methods* **8**, 879-886 (2011).
46. Schumacher, R.J. et al. ATP-dependent chaperoning activity of reticulocyte lysate. *J. Biol. Chem.* **269**, 9493-9499 (1994).
47. Holmberg, C.I., Staniszewski, K.E., Mensah, K.N., Matouschek, A. & Morimoto, R.I. Inefficient degradation of truncated polyglutamine proteins by the proteasome. *EMBO J.* **23**, 4307-4318 (2004).
48. Ramdzan, Y.M. et al. Tracking protein aggregation and mislocalization in cells with flow cytometry. *Nat. Methods* **9**, 467-U476 (2012).

Chapter 3: Multidimensional chemical control of CRISPR-Cas9

This chapter is adapted from the following manuscript:
Moore, C.L.[1], Maji, B.[1], Zetsche, B., Volz, S.E., Zhang, F., Choudhary, A.[2], Shoulders, M.D.[2]
Nature Chemical Biology, **2017,** 13, pp 9–11. [1]Equal Contributions; [2]Co-Corresponding Authors.

3.1 Author Contributions

C.L.M., B.M., B.Z., M.D.S., and A.C. planned research and analyzed data; B.M., C.L.M., B.Z., F.Z., M.D.S., and A.C. designed experiments; B.M., C.L.M. performed all transcriptional control and Western blot experiments. B.M. and C.L.M. performed Cas9 gene editing fluorescent reporter knockout experiments. B.M. performed NGS analysis of Cas9 gene editing with assistance from C.L.M. and B.Z. B.M. performed surveyor assay experiments with assistance from B.Z. C.L.M., B.Z., and S.E.V. cloned plasmids used in experiments; M.D.S. and A.C. supervised research.

3.2 Abstract

Cas9-based technologies have transformed genome engineering and the interrogation of genomic functions, but methods to control such technologies across numerous dimensions, including dose, time, specificity, and mutually exclusive modulation of multiple genes, are still lacking. We conferred such multi-dimensional controls on diverse Cas9 systems by leveraging small molecule-regulated protein degron domains. Application of our strategy to both Cas9-mediated genome editing and transcriptional activities opens new avenues for systematic genome interrogation.

3.3 Introduction

RNA-guided endonucleases, such as Cas9, are easily targeted to any genomic locus using single guide RNAs (sgRNAs)[1,2], ushering in a slew of transformative technologies. For example, Cas9 enables facile genomic alterations, as well as robust self-propagation of such alterations throughout a species population via gene drives[3]. Furthermore, catalytically inactive Cas9 (dCas9) can be fused to a wide range of effectors, including fluorescent proteins for genome imaging[4], enzymes that modify DNA or histones for epigenome editing[5], and transcription-regulating domains for controlling endogenous gene expression[6].

Despite such advances, a critical need still exists for methods to precisely regulate Cas9 activities across multiple dimensions, including dose, target, and time[7]. Finely-tuned control of Cas9 levels is important[8], as high Cas9 concentrations result in elevated off-target genomic alterations. Rapid disabling of activity after a desired genomic modification is also valuable to prevent off-target activity[9]. In the context of gene regulation by dCas9-based transcriptional activators, dose control of transcript expression levels is essential to permit induction of physiologically relevant levels of mRNA transcripts. The abilities to rapidly reverse transcript induction and to control the expression of multiple transcripts independently are also highly desirable. Ideally, methods that provide such controls should be readily adaptable to diverse RNA-guided nucleases that continue to rapidly emerge.

We sought small molecule-regulated systems that would fulfill these needs. Previously, we applied destabilizing domains (DDs)[10-12] to confer similar controls on transcription factors[13,14]. Briefly, structurally unstable protein domains derived from *E. coli* dihydrofolate reductase (DHFR)[10] or the estrogen receptor (ER50)[11] are fused to a transcription factor. These largely unfolded domains target the fusion protein for rapid proteasomal degradation. Small molecules that can bind and stabilize the poorly populated folded state of the DDs prevent proteasomal degradation of the fusion protein in a concentration-dependent manner, allowing transcription factor function[13,14]. We envisioned that such small molecule-regulated DDs could be similarly deployed to establish Cas9 systems with multi-dimensional control of genome editing and transcriptional activities.

3.4 Results and Discussion

3.4.1 Developing a Potent, Small Molecule-Regulated, and Cas9-Mediated Transcriptional Control System

We began by genetically linking the DHFR DD to the N-terminus of catalytically inactive *Streptococcus pyogenes* Cas9 (dSpCas9) fused to VP16-based transcriptional activation

domains (e.g., VP64 or VP192)[15]. Addition of trimethoprim (TMP), a DHFR-stabilizing small molecule, to cells transfected with either DHFR.dSpCas9.VP64 or a previously reported DHFR.dSpCas9.VP192 fusion[16] and appropriate sgRNAs yielded minimal induction of target genes or displayed high basal activity (**Figure 3.1 and Figure 3.2**, respectively). These findings were expected, as dSpCas9.VP64 alone is insufficient for inducing transcription of many target genes[17].

We shifted our focus to a second-generation Cas9-based transcription system[17], in which sgRNAs bear an RNA aptamer that recruits transcription activation domains (e.g., PP7.VP64[18]) to dSpCas9. We conferred conditional activity on this system by fusing the DHFR DD to the PP7.VP64 transcription activation domain (**Figure 3.3**). Cells transiently expressing DHFR.PP7.VP64, dSpCas9, and a sgRNA targeting *IL1RN* showed robust upregulation of *IL1RN* mRNA upon treatment with TMP, demonstrating chemogenic control of transcript induction (**Figure 3.4**). Further analysis of mRNA levels in this cell population at different treatment times revealed that IL1RN mRNA levels started to increase within 8 hours of small molecule treatment (**Figure 3.5**). Importantly, we observed minimal basal transcriptional activation in vehicle-treated samples.

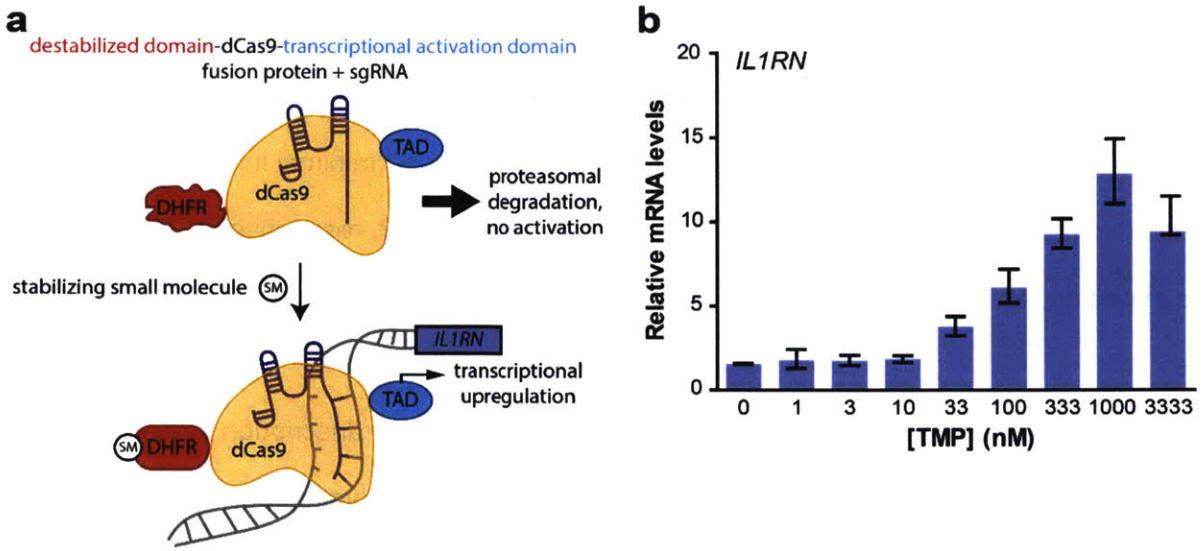

Figure 3.1 | Regulation of dSpCas9.VP64 transcriptional activity using directly fused destabilized domain to dSpCas9

(**a**) Model for small molecule-regulated transcriptional activation mediated by DHFR.dSpCas9.VP64.

(**b**) qPCR analysis of HEK293T cells transiently transfected with DHFR.dSpCas9.VP64 and four requisite sgRNAs targeting *IL1RN* following treatment with increasing concentrations of TMP for 18 h prior to qPCR analysis. Error bars represent standard deviation across four technical replicates.

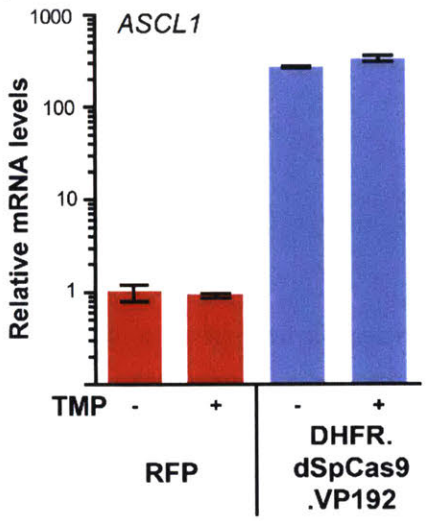

Figure 3.2 | DHFR.dSpCas9.VP192 possesses transcriptional activity prior to small molecule induction

qPCR analysis of HEK293T cells expressing DHFR.dSpCas9.VP192 and requisite sgRNAs targeting *ASCL1*. Transiently transfected cells were treated with 10 μM TMP for 18 h prior to qPCR analysis. Error bars represent standard deviation across four technical replicates.

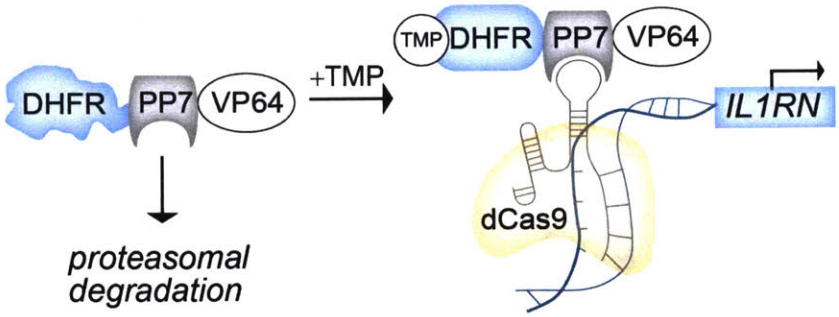

Figure 3.3 | Destabilizing domain regulation of an accessory protein in a second generation dCas9 transcriptional system relies on proteasomal degradation.

Schematic illustrating how small molecule-mediated regulation of a second generation dCas9 transcriptional system is achieved. In this approach, a destabilizing domain based on dihydrofolate reductase of *E. coli* (DHFR) is genetically fused to an accessory protein consisting of an RNA-binding protein heavily used in systems biology (phage coat protein PP7) and a transcriptional activation domain (VP64). Upon expression in a living cell, DHFR is largely in a conformation recognized for proteasomal degradation, resulting in turnover of the fused PP7.VP64 domains. In the presence of a stabilizing small molecule (trimethoprim, TMP), DHFR remains in a stable conformation and is not recognized for degradation by the proteasome, resulting in realized expression of DHFR.PP7.VP64 fusion and functional downstream activity of those domains. In this example, PP7 binds to a cognate RNA aptamer present in a loop of an sgRNA presented by a sgRNA-dCas9 ribonucleoprotein complex targeted to a region upstream of endogenous *IL1RN* in human cells.

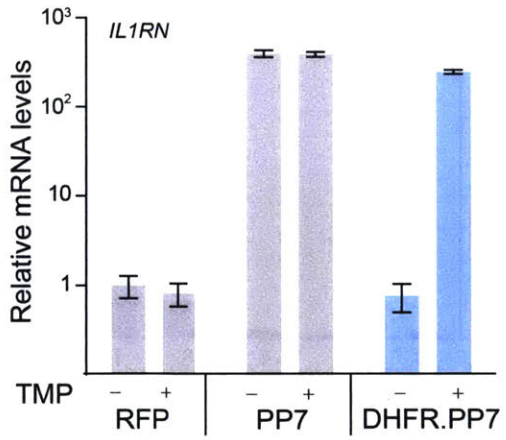

Figure 3.4 | Small molecule regulation of a robust dSpCas9-mediated transcriptional activation system requires fusion of destabilizing domain to accessory protein

Small molecule-mediated transcription induction via a destabilized domain-fused transcription activation domain (DHFR.PP7.VP64), dSpCas9, and a sgRNA (top). HEK293T cells transfected with dSpCas9 and either an RFP control, PP7.VP64, or trimethoprim (TMP)-regulated DHFR.PP7.VP64 targeted to *IL1RN* were treated with 10 μM TMP for 18 h prior to qPCR analysis (bottom). Error bars represent SEM from biological replicates (n = 3).

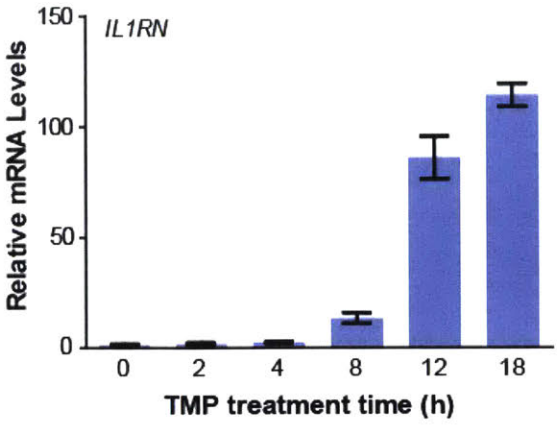

Figure 3.5 | Temporal dynamics of transcriptional upregulation with DHFR.PP7.VP64 and dSpCas9.

qPCR analysis of HEK293T cells expressing dSpCas9, DHFR.PP7.VP64 and sgRNAs targeted to *IL1RN* following treatment with TMP (100 nM) for increasing amounts of time up to 18 h. Error bars represent standard deviation across four technical replicates.

3.4.2 Advantages of New DD-dCas9 Transcriptional Systems Include Rapid Reversibility

Existing small molecule-regulated, dCas9 transcriptional activators leave transcription continuously on[7,19], unlike in biological systems where both transcription turn-on *and* turn-off are often strictly regulated. An advantage of our approach is the potential for turning-off transcription upon removal of the DD-stabilizing small molecule. To test this idea, we treated cells transfected with DHFR.PP7.VP64, dSpCas9, and sgRNA with TMP to upregulate endogenous genes prior to a chase period with media lacking TMP (**Figure 3.6**, top). Removal of TMP resulted in rapid depletion of induced mRNA transcript levels within <8 h for *IL1RN* and *NANOG* (**Figure 3.6**). For IL1RN, we validated the resulting rapid depletion in replicate time course experiments (**Figure 3.7**). Thus, our system permits user-defined turn-on and turn-off of endogenous gene transcription.

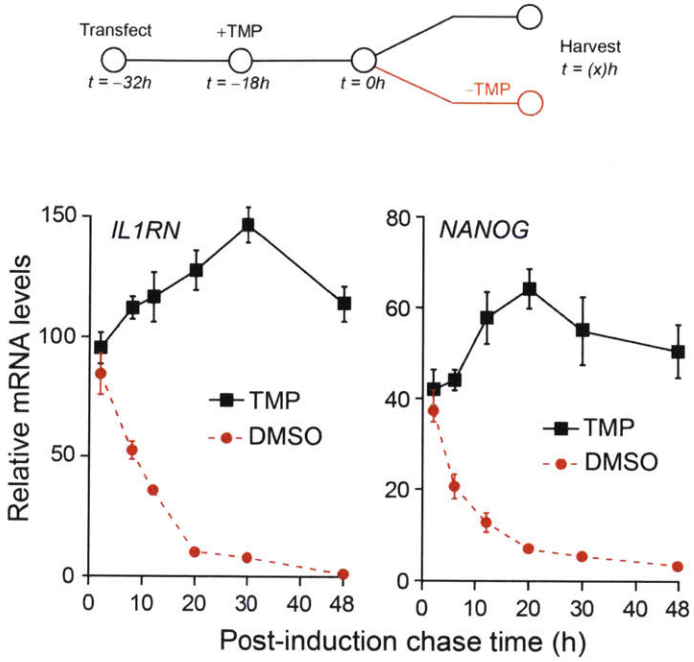

Figure 3.6 | Rapid turn-off of transcriptional upregulation with DHFR.PP7.VP64 and dSpCas9

Cells were transfected and treated with 100 nM TMP to upregulate endogenous *IL1RN* (additional biological replicates shown in **Figure 3.7**) or *NANOG*. After 18 h of TMP treatment, cells were provided with fresh media containing or lacking TMP prior to harvesting and analysis by qPCR. Error bars represent S.D. from technical replicates (n = 4).

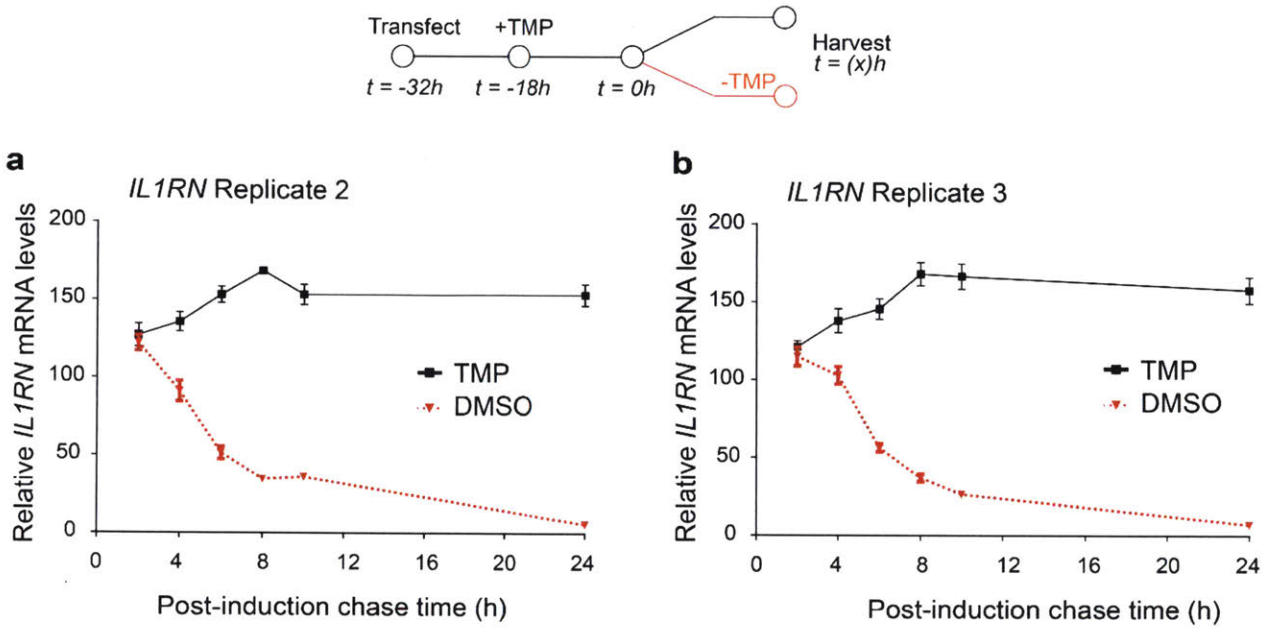

Figure 3.7 | Validation of reversibility of transcriptional upregulation with DHFR.PP7.VP64 and dSpCas9 targeted to *IL1RN*

Time course of the reversible transcriptional induction experiment utilizind DHFR. PP7.VP6 system (above). Two additional biological replicates (below, **a** and **b**) of the experiment shown in **Figure 3.6** for reversible transcriptional induction of *IL1RN*. Error bars represent standard deviation across four technical replicates.

3.4.3 DD-dCas9 Transcriptional Systems Enable Orthogonal Gene Regulation in a Single Population

Another attractive feature of our strategy is that several orthogonal small molecule–degron pairs exist[10,11], theoretically enabling independent control of the transcription of multiple genes. For example, the ER50 DD stabilized by the small molecule (*Z*)-4-hydroxytamoxifen (4OHT)[11] is orthogonal to the TMP–DHFR pair. Conveniently, MS2.p65.HSF1 is a transcription activation domain whose cognate RNA aptamer is distinct from that bound by the PP7.VP64 domain, and therefore is functionally orthogonal to PP7.VP64[17,18]. Similar to our previous observations using DHFR.PP7.VP64 and TMP, expression of ER50.MS2.p65.HSF1 with dSpCas9 and an appropriate sgRNA yielded 4OHT-dependent transcript induction (**Fig. 3.8**). Immunoblotting confirmed that this induction was accompanied by a 4OHT-dependent increase in ER50.MS2.p65.HSF1 protein levels (**Figure 3.9**). As expected, we observed a similar increase in ER50.MS2.p65.HSF1 protein levels upon proteasome inhibition using MG-132, confirming that 4OHT rescues the transcription activation domain from degradation (**Figure 3.9**). Most importantly, co-expression of both DHFR.PP7.VP64 and ER50.MS2.p65.HSF1, along with dSpCas9 and appropriate sgRNAs, permitted conditional and orthogonal activation of multiple endogenous genes when either TMP, 4OHT, or both TMP and 4OHT were added to cells (**Figure 3.8**).

As the discovery and development of new and improved RNA-guided nucleases is currently quite rapid, an ideal small molecule-regulated method must easily translate to new versions of RNA-guided nucleases with minimal engineering. We note that the ER50 DD can also be easily deployed to regulate the activity of catalytically inactive *Staphylococcus aureus* Cas9 (dSaCas9)[20], a Cas9 variant from a different species than SpCas9 with distinct sgRNAs (**Figure 3.10**). While this result highlights the ready adaptability of our approach to next-generation RNA-guided, dCas9-based transcriptional activators, the implementation of dSaCas9

expands the current orthogonality of small molecule gene regulation by immediately providing an alternate tool to dSpCas9.

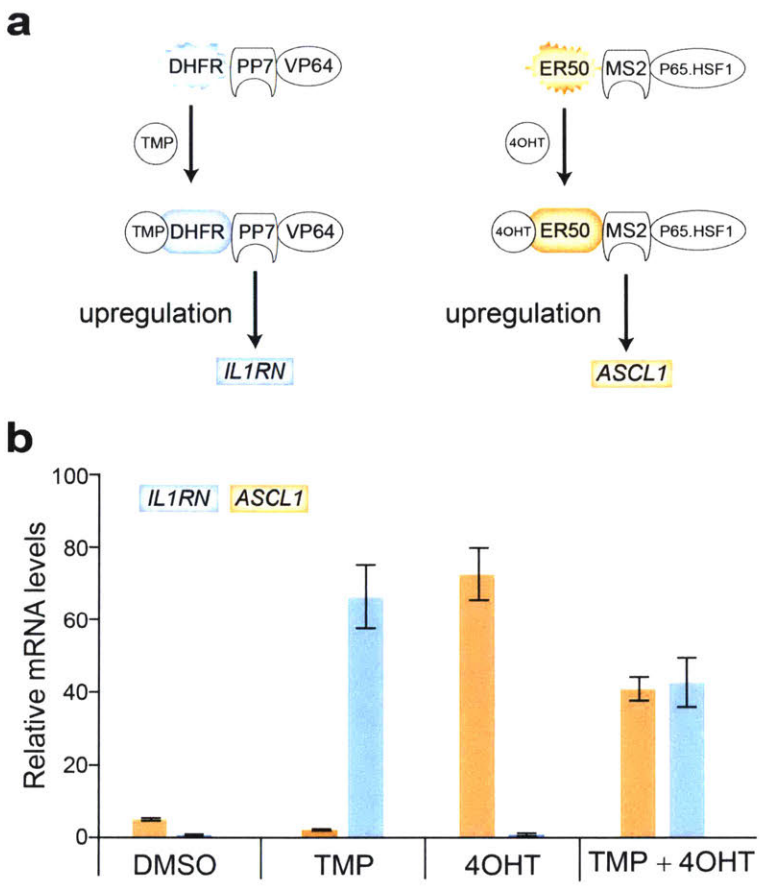

Figure 3.8 | Independent, small molecule-mediated control of transcript levels for two genes

(a) Cells were transfected with dSpCas9 and two orthogonal destabilized domain-regulated transcription activation systems targeted to separate genes. DHFR.PP7.VP64 was targeted to *IL1RN* and ER50.MS2.p65.HSF1 was targeted to *ASCL1*.

(b) Transfected cells were treated as indicated with 100 nM TMP and/or 10 nM 4OHT for 18 h prior to qPCR analysis. Error bars represent SEM from biological replicates (n = 3).

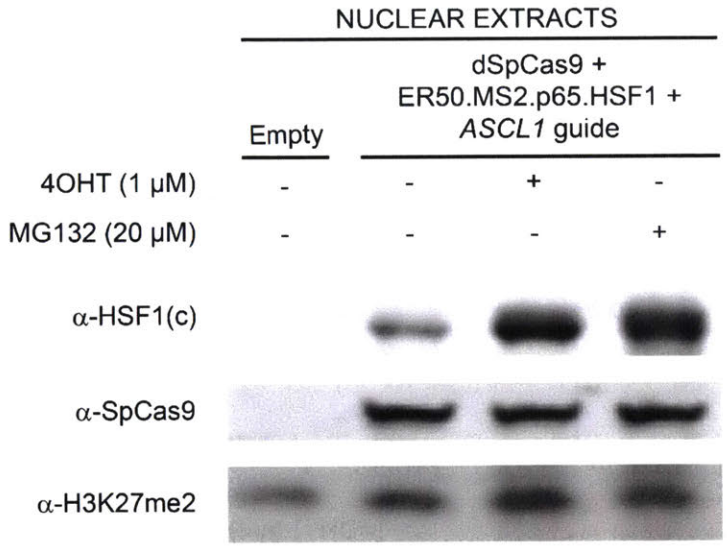

Figure 3.9 | Western blot analysis of destabilized domain-regulated transcriptional activation effector in the absence and presence of stabilizing small molecule

Western blot analysis of HEK293T cells transfected with dSpCas9 and 4OHT-regulated ER50.MS2.p65.HSF1 targeted to *ASCL1*. Treatment with stabilizing small molecule 4OHT (1 μM, 24 h) or proteasome inhibitor MG132 (10 μM, 12 h) correlates with the substantial increase of a band corresponding to ER50.MS2.p65.HSF1 (detected using a C-terminal antibody against HSF1; α-HSF1(c)), while no changes are observed in dSpCas9 levels.

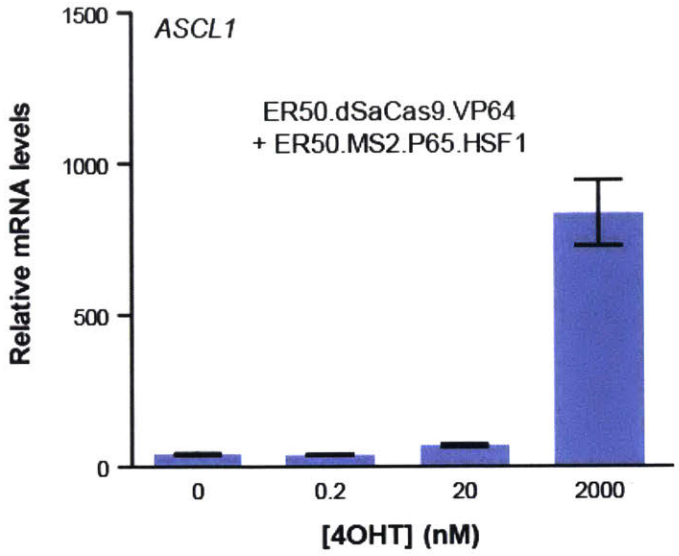

Figure 3.10 | Regulation of dSaCas9 transcriptional activity using the ER50 DD

qPCR analysis of HEK293T cells expressing ER50.dSaCas9.VP64, ER50.MS2.P65.HSF1 and sgRNAs targeted to *ASCL1* following treatment with increasing concentrations of 4OHT for 18 h. Error bars represent standard deviation across four technical replicates.

3.4.4 DD-dCas9 Transcriptional Control Systems Possess a Broad, Robust Range of Dosability

Meaningful interrogation of gene function frequently requires the ability to dose transcript levels across a wide dynamic range. The pharmacologic chaperoning-based mechanism of DD stabilization can engender a broad range of dosable regulation[13,14]. Indeed, the levels of TMP-mediated *IL1RN* mRNA induction using the DHFR-regulated PP7.VP64 transcription activation domain can be controlled across orders of magnitude simply by modulating the dose of TMP (**Figure 3.11a**). *ASCL1* mRNA upregulation can also be controlled across a similar range by modulating the 4OHT dose in cells expressing the ER50.MS2.p65.HSF1 transcription activation domain (**Figure 3.11b**). This precise tuning of transcriptional activity demands little optimization of dCas9, sgRNA, or transcription activation domain expression levels, simply requiring the addition of an appropriate concentration of TMP or 4OHT.

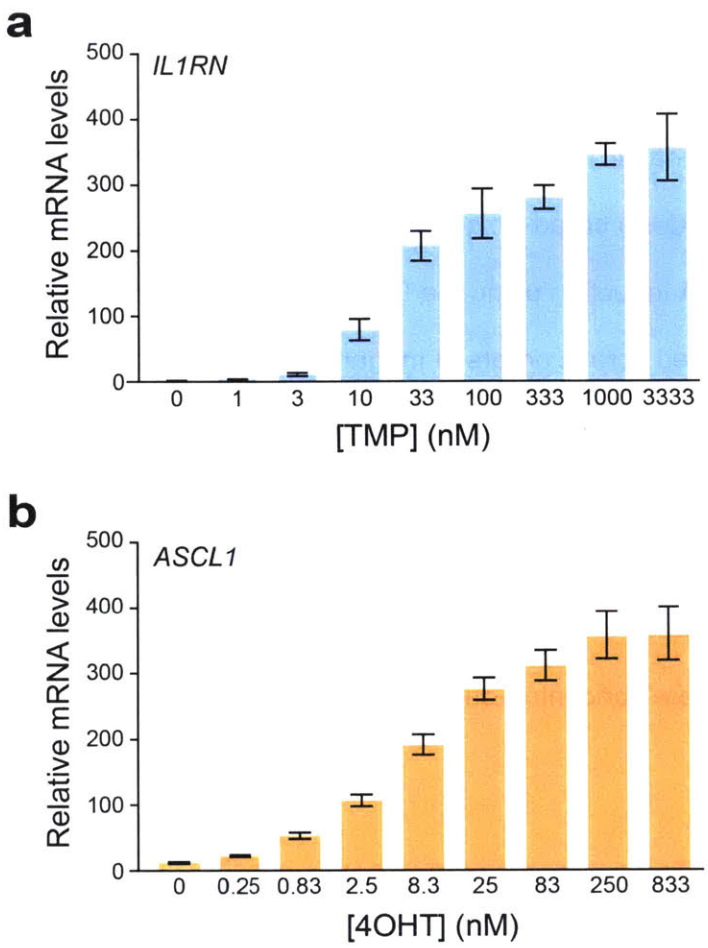

Figure 3.11 | Destabilized domain-regulated transcriptional activation effectors achieve highly dose-responsive endogenous gene upregulation

Cells were transfected with dSpCas9, appropriate sgRNAs, and either DHFR.PP7.VP64 (left) or ER50.MS2.p65.HSF1 (right) targeted to *IL1RN* or *ASCL1*, respectively. Transfected cells were treated with increasing concentrations of TMP or 4OHT for 18 h prior to qPCR analysis. Error bars represent S.D. across technical replicates (n = 4).

3.4.5 Robust Regulation of Catalytic Cas9 Can Be Achieved with Destabilizing Domains

These data demonstrate chemogenic control of endogenous gene transcription across several dimensions, including dose, time, and orthogonal modulation of multiple genes. Next, we sought to engineer catalytically active Cas9 variants displaying dose and temporal control of nuclease activity and specificity. Such systems would enable regulated and specific genome editing. Temporal control of nuclease activity may also find use in emerging technologies, including gene drives[3,21]. Consistent with our observations for dSpCas9.VP64 (**Figure 3.1**), we found that the fusion of a DHFR or ER50 DD to either the N- or C-terminus of catalytically active SpCas9 resulted in limited small molecule control (**Figure 3.12**). In contrast, fusing DHFR or ER50 DDs to both the N- and C-termini of SpCas9 (e.g., DHFR.SpCas9.DHFR) was more successful (**Figure 3.12**), providing inducible control of nuclease activity. Immunoblotting illustrates the TMP- or 4OHT-dependent increases in DHFR.SpCas9.DHFR and ER50.SpCas9.ER50 protein levels that accompanied this activity induction (**Figure 3.13**). Furthermore, we observed strongly dose-dependent control of gene editing activity for both our DD.SpCas9.DD constructs (**Figure 3.14**), with basal activity similar to that previously reported for a split SpCas9 regulated by intein self-splicing[22] (**Figure 3.15**).

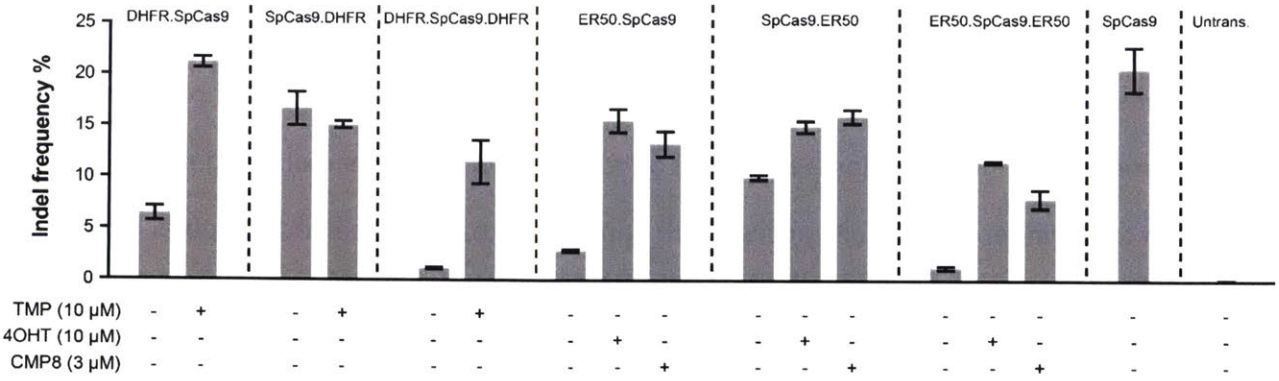

Figure 3.12 | Regulation of genome editing with diverse destabilizind domain SpCas9 construct architectures

Indel frequency of assorted Cas9 constructs in HEK293T cells either in the presence or absence of stabilizing small molecules (TMP, 4OHT, and CMP8). Samples were prepared by transfecting HEK293T cells with the indicated plasmid along with *EMX1(1)*-targeted sgRNA and incubated for 72 h either in the presence or absence of small molecules. Both TMP and 4OHT were used at 10 μM while CMP8 was used at 3 μM. Indel-frequencies were measured by next-generation sequencing. The studies revealed that dual DD-regulated Cas9 systems provide superior regulation of gene editing ability over the corresponding single DD-regulated analogs. Error bars for each panel represent standard deviation from biological replicates (n = 4).

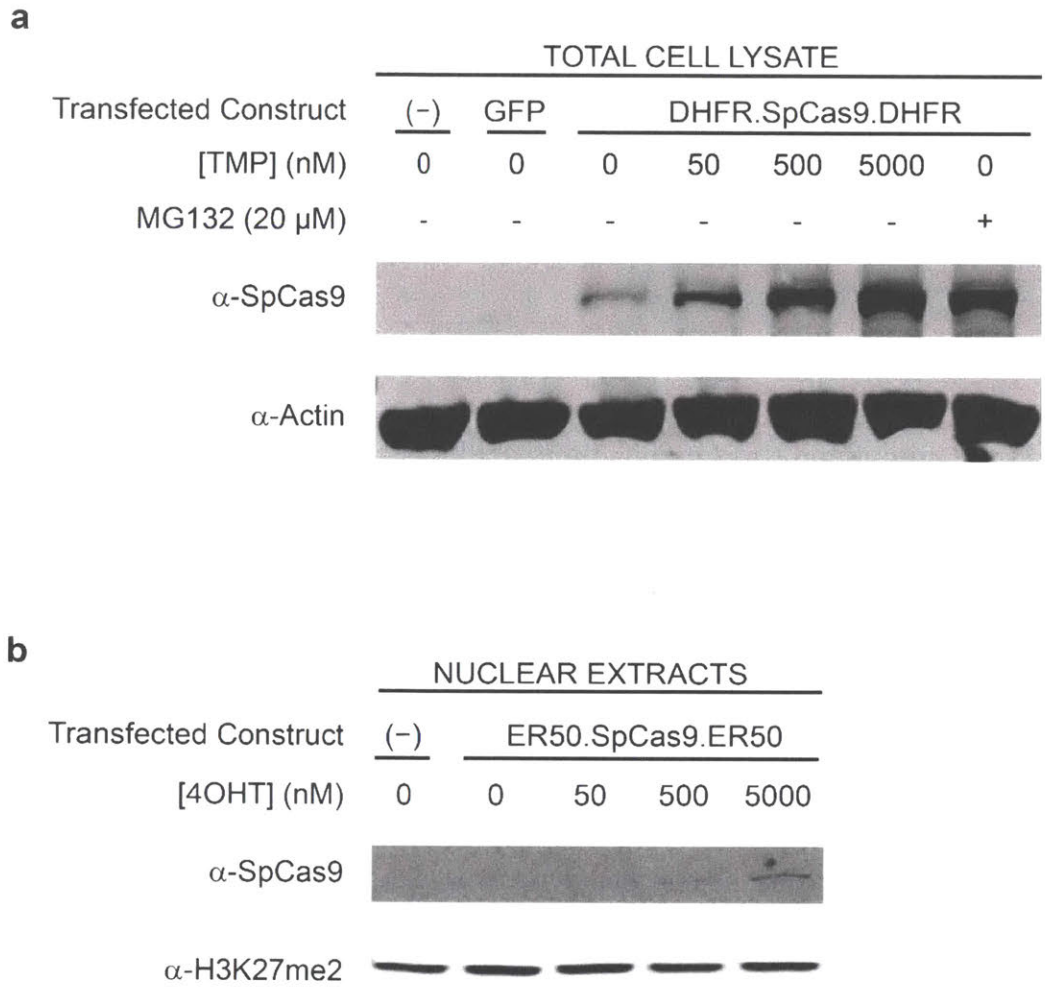

Figure 3.13 | Western blot analysis of DD-fused SpCas9 protein levels in the absence and presence of the stabilizing small molecules TMP or 4OHT.

(**a**) Western blot analysis of total cell lysates obtained from HEK293T cells transfected with DHFR.SpCas9.DHFR and incubated in the presence or absence of increasing amounts of TMP (50–5000 nM) for 24 h or the proteasome inhibitor MG-132 for 12 h.

(**b**) Western blot analysis of nuclear extracts obtained from HEK293T cells transfected with ER50.SpCas9.ER50 and incubated in the presence or absence of increasing amounts of 4OHT (10–1000 nM) for 24 h.

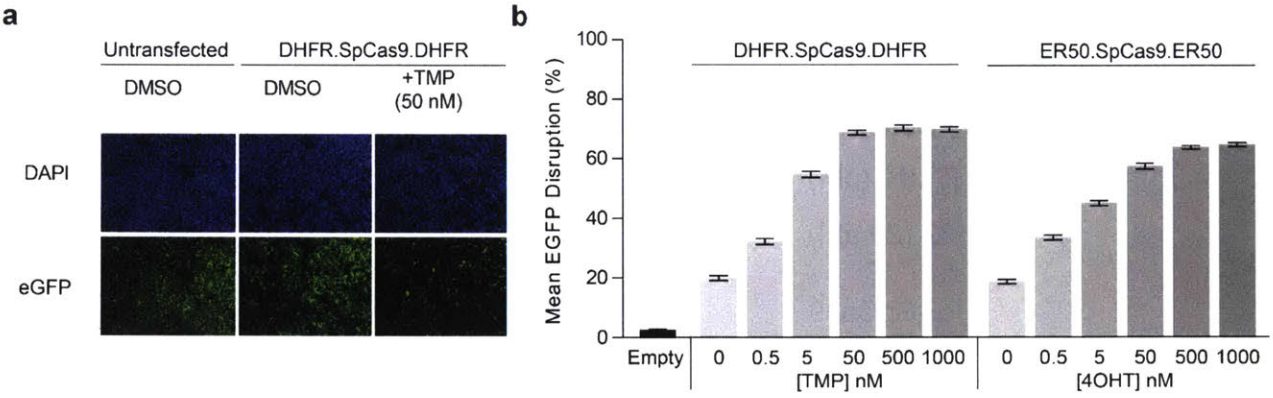

Figure 3.14 | DD-regulated Cas9 gene editing activity is highly dose-responsive

(**a**) Representative images of conditional control of Cas9-mediated eGFP knockout in U2OS.eGFP-PEST cells. Cells nucleofected with DHFR.SpCas9.DHFR show a marked decrease in eGFP signal detected with a high-content imaging microscope after treatment with the DD-stabilizing small molecule TMP (50 nM) for 48 h. Error bars represent standard deviation from biological replicates (n = 5).

(**b**) High dosability of genome editing observed in U2OS.eGFP-PEST cells nucleofected with DHFR.SpCas9.DHFR or ER50.SpCas9.ER50 following treatment with increasing concentrations of the stabilizing small molecules TMP or 4OHT (0.5–1000 nM) for 48 h. Error bars represent standard deviation from biological replicates (n = 5).

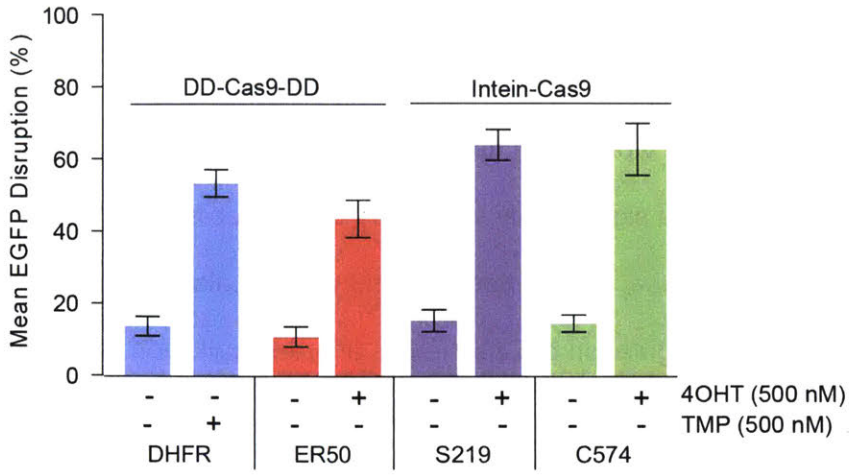

Figure 3.15 | Assessment of small molecule-regulated Cas9 gene editing systems

Quantitation of basal and induced genome editing activities of Cas9 systems mediated by DDs or a 4OHT-sensitive self-splicing intein (S219 and C574). Measurements were performed using the eGFP disruption assay from **Figure 3.14a** in U2OS.eGFP-PEST cells after 48 h of treatment with 500 nM TMP or 4OHT. Error bars represent SEM from biological replicates (n = 5).

3.4.6. Dosability of DD-Cas9 Enables Optimization of Gene Editing Targeting Efficiency

Off-target activity of Cas9 nucleases can lead to catastrophic biological events, including chromosomal translocations[23]. Controlling genome editing specificity is therefore of much interest. Truncation of sgRNAs significantly enhances Cas9 specificity, and employing nickase variants, a FokI–dCas9 nuclease, or high fidelity Cas9 variants also offer improvements[23]. We anticipated that the precise regulation of Cas9 levels afforded by fusion to small molecule-controlled DDs would allow us to titer in optimal Cas9 concentrations to maximize on-target while minimizing off-target gene editing[22,24]. Indeed, we observed enhanced specificity for on-target versus known off-target sites of *VEGFA* and *EMX1* upon administering optimized doses of TMP or 4OHT for the DHFR.SpCas9.DHFR and the ER50.SpCas9.ER50 systems, respectively (summary in **Figure 3.16a–d**, with extensive data in **Table 3.1**, **Figure 3.17** and **Figure 3.18**).

Limiting Cas9 activity to a short temporal window is another promising avenue to enhance genome-editing specificity, and has been accomplished previously by delivering a ribonucleoprotein complex of Cas9 and sgRNA[25]. Using DD-regulated Cas9, we should be able to further control the size of the temporal window (as demonstrated in **Figure 3.5**, **Figure 3.6** and **Figure 3.7** for DD-regulated transcription activation domains). In the context of gene editing, this feature is illustrated by our observation that a short pulse of TMP (6 h) in U2OS.eGFP-PEST cells co-expressing DHFR.SpCas9.DHFR and eGFP-targeting sgRNAs resulted in significantly less eGFP knockout than a 48 h pulse over the same timecourse (**Figure 3.19**). Cumulatively, these results highlight the potential of DD-regulation to control the specificity of Cas9 gene editing tools. We note that it is straightforward to extend our modular method to other next-generation RNA-guided endonucleases (see, for example, DD-mediated control of SaCas9 activity[20] in **Figure 3.20**).

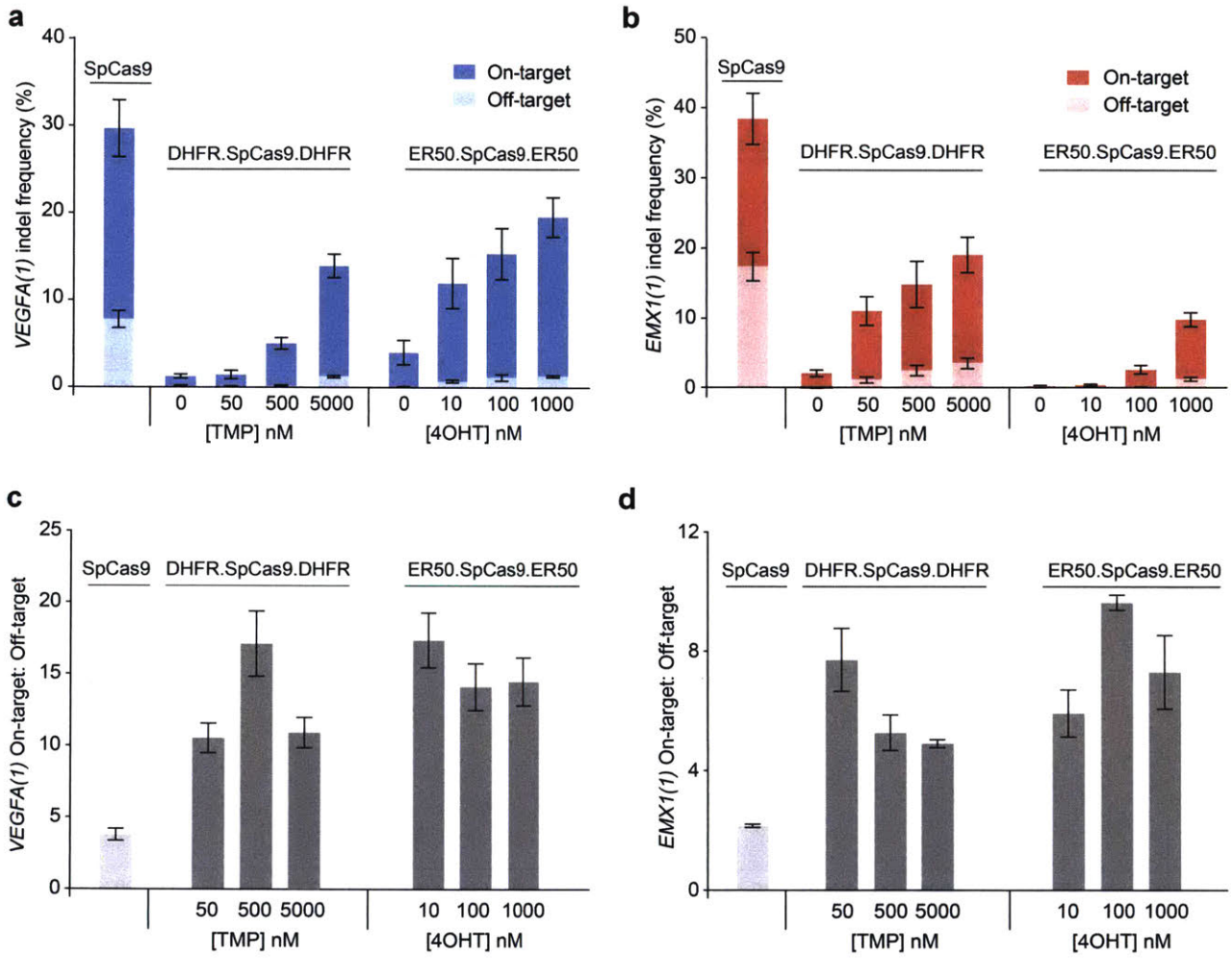

Figure 3.15 | Assessment of targeting efficiency in small molecule-regulated Cas9 gene editing systems

Indel frequency of assorted Cas9 constructs in HEK293T cells either in the presence or absence of stabilizing small molecules (TMP, 4OHT, and CMP8). Samples were prepared by transfecting HEK293T cells with the indicated plasmid along with *EMX1(1)*-targeted sgRNA and incubated for 72 h either in the presence or absence of small molecules. Both TMP and 4OHT were used at 10 μM while CMP8 was used at 3 μM. Indel-frequencies were measured by next-generation sequencing. The studies revealed that dual DD-regulated Cas9 systems provide superior regulation of gene editing ability over the corresponding single DD-regulated analogs. Error bars for each panel represent standard deviation from biological replicates (n = 4).

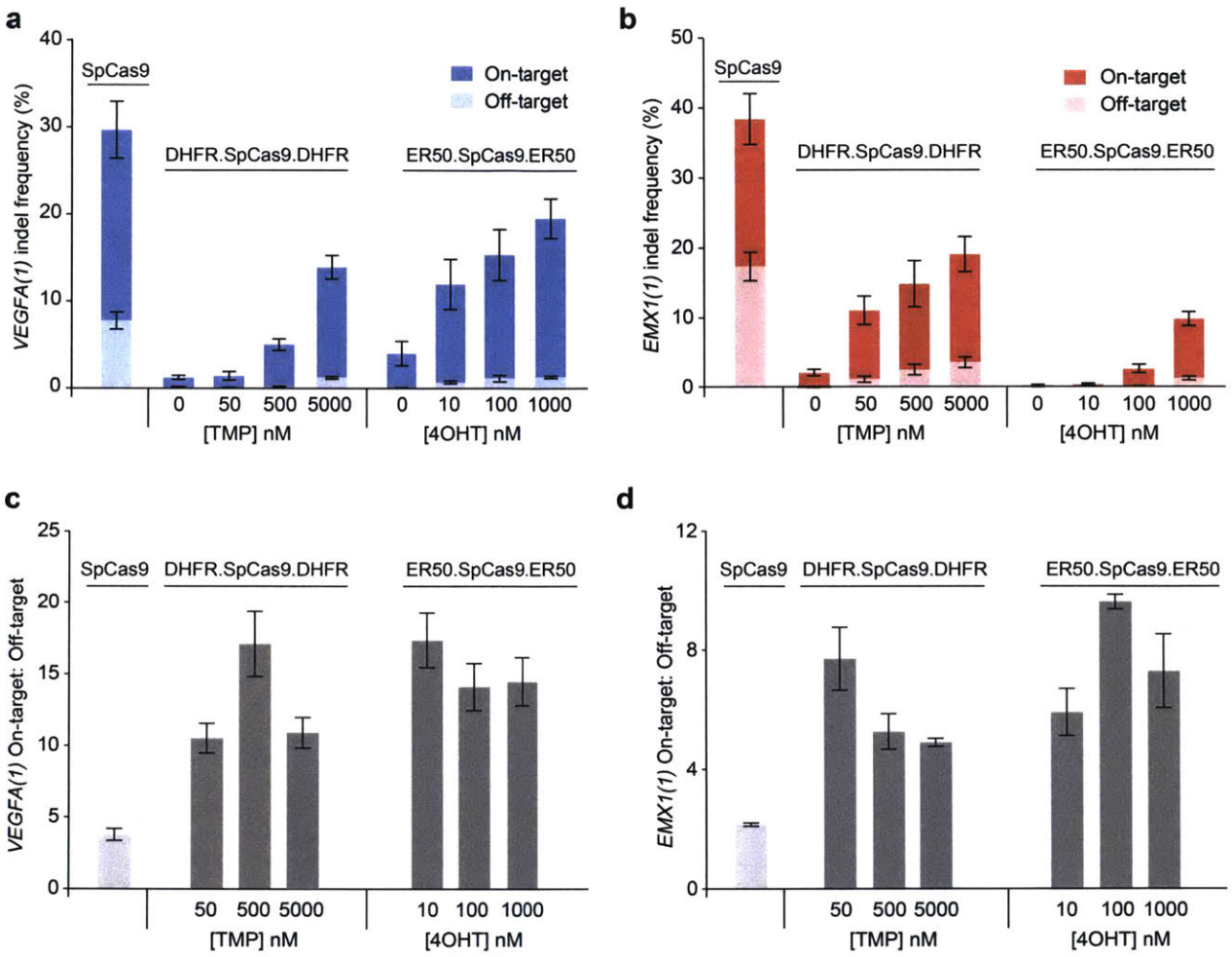

Figure 3.16 | On-target editing efficiency of DD-SpCas9 gene editing systems is a tunable parameter

(a–b) TMP- and 4OHT-dose-dependent control of on- and off-target activity of DD.SpCas9.DD targeting *VEGFA* **(a)** or *EMX1* **(b)**. HEK293T cells were transfected with SpCas9, DHFR.SpCas9.DHFR, or ER50.SpCas9.ER50 and treated with the indicated doses of vehicle, TMP, or 4OHT for 48 h prior to genomic DNA extraction and analysis of on-target and off-target indel frequencies by next-generation sequencing. Error bars represent SEM from biological replicates (n = 4).

(c–d) Ratiometric representation of On-target:Off-target indel frequencies of DD.SpCas9.DD for *VEGFA* **(c)** and *EMX1* **(d)**. Error bars represent SEM from biological replicates (n = 4).

Table 3.1 | Average indel frequency assessed by next-generation sequencing

Sample	ON	OT1	OT2	OT3	OT4	OT5
GFP_*EMX1(1)*	0.16	0.14	0.12	0.08	0.03	0.10
SpCas9+*EMX1(1)*	41.10	18.85	11.50	1.37	0.02	0.41
DHFR.SpCas9.DHFR+*EMX1(1)*	2.23	0.18	0.15	0.07	0.03	0.08
+TMP 50 nM	11.67	1.55	0.56	0.18	0.03	0.11
+TMP 500 nM	16.33	3.15	1.23	0.22	0.04	0.15
+TMP 5000 nM	19.76	4.02	1.47	0.23	0.02	0.23

Sample	ON	OT1	OT2	OT3	OT4	OT5
GFP_*EMX1(1)*	0.17	0.20	0.16	0.12	0.05	0.13
SpCas9+*EMX1(1)*	43.22	20.80	12.41	1.64	0.06	0.44
ER50.SpCas9.ER50+*EMX1(1)*	0.56	0.31	0.15	0.08	0.02	0.20
+4OHT 10 nM	0.72	0.12	0.14	0.11	0.07	0.23
+4OHT 100 nM	4.02	0.18	0.18	0.07	0.06	0.20
+4OHT 1000 nM	14.04	2.10	0.72	0.12	0.04	0.32

Sample	ON	OT1
GFP_*VEGFA(1)*	0.16	0.14
SpCas9+*VEGFA(1)*	29.78	7.83
DHFR.SpCas9.DHFR+*VEGFA(1)*	1.29	0.14
+TMP 50 nM	1.58	0.19
+TMP 500 nM	5.09	0.31
+TMP 5000 nM	14.00	1.28

Sample	ON	OT1
GFP_*VEGFA(1)*	1.81	0.45
SpCas9+*VEGFA(1)*	29.25	6.38
ER50.SpCas9.ER50+*VEGFA(1)*	4.11	0.21
+4OHT 10 nM	12.04	0.77
+4OHT 100 nM	15.43	1.18
+4OHT 1000 nM	19.66	1.36

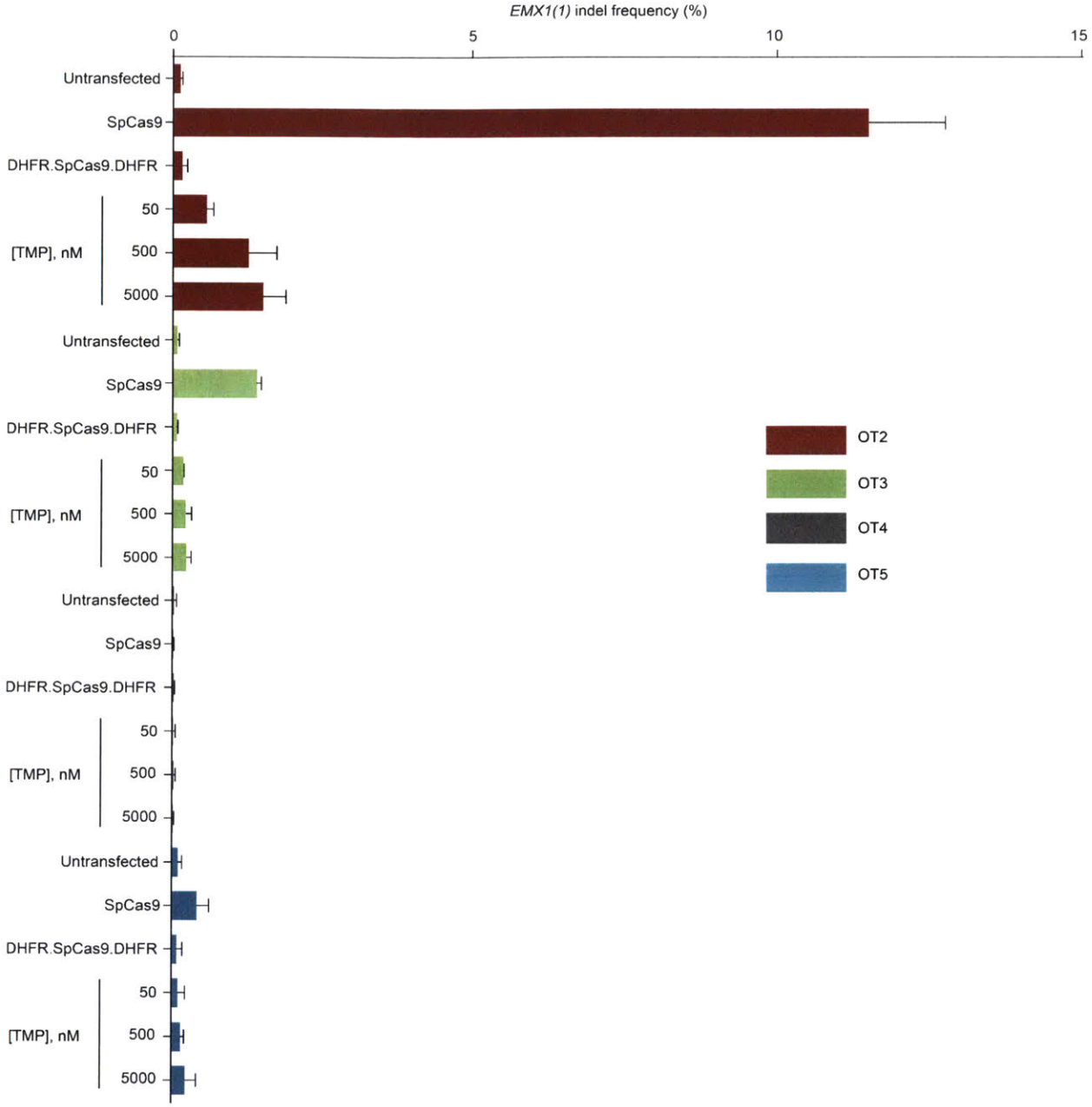

Figure 3.17 | Specificity of DHFR.SpCas9.DHFR gene editing system

Indel frequency plots determined by next-generation sequencing to measure off-target activities for DHFR.SpCas9.DHFR construct targeted to the *EMX1* gene. Samples were prepared by transfecting HEK293T cells with the DHFR.SpCas9.DHFR plasmid along with *EMX1(1)* targeted sgRNA and incubated for 72 h either in the presence or absence of increasing concentrations of TMP (50–5000 nM). Error bars for each panel represent standard deviation from biological replicates (n = 4).

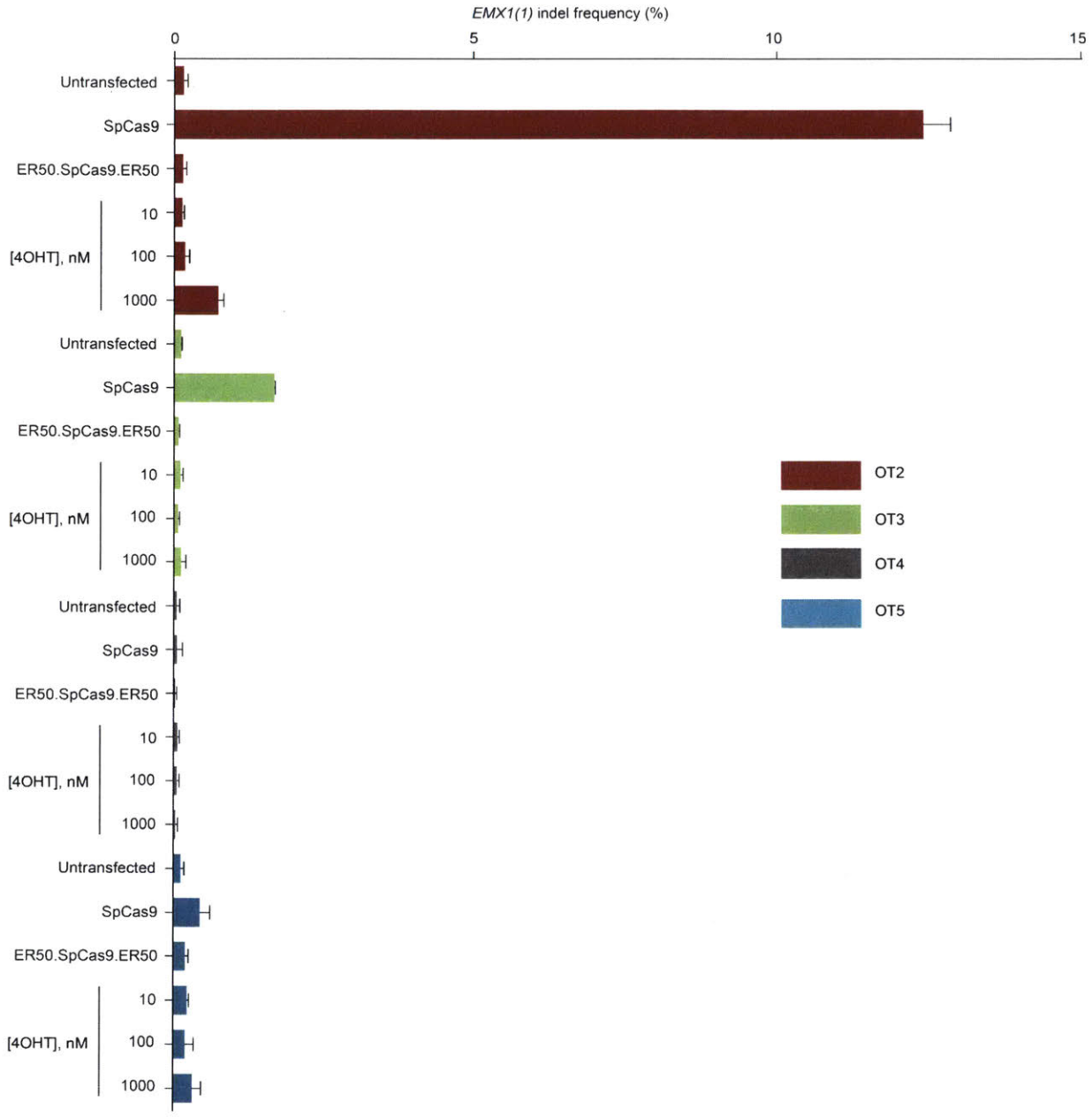

Figure 3.18 | Specificity of ER50.SpCas9.ER50 gene editing system.

Indel frequency plots determined by next-generation sequencing to measure off-target activities for the ER50.SpCas9.ER50 construct targeted to the *EMX1* gene. Samples were prepared by transfecting HEK293T cells with the ER50.SpCas9.ER50 plasmid along with *EMX1(1)* targeted sgRNA and incubated for 72 h either in the presence or absence of increasing concentrations of 4OHT (10–1000 nM). Error bars for each panel represent standard deviation from biological replicates (n = 4).

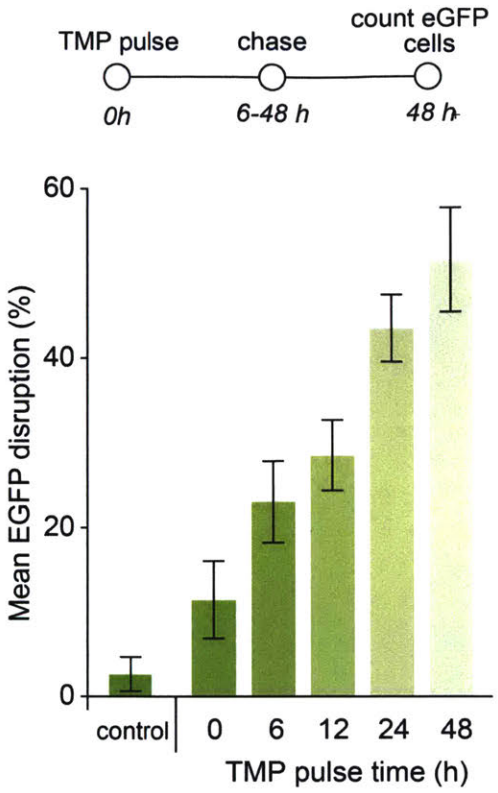

Figure 3.19 | Temporal control of DHFR.SpCas9.DHFR-mediated genome editing analyzed by an eGFP disruption assay

U2OS.eGFP-PEST cells nucleofected with a plasmid expressing DHFR.SpCas9.DHFR and a sgRNA targeting eGFP were incubated with the indicated concentrations of TMP for increasing periods of time (6–48 h) prior to media swap to remove TMP. $eGFP^+$ cells were counted using automated, high-content imaging microscopy. Error bars represent SEM from biological replicates (n = 5).

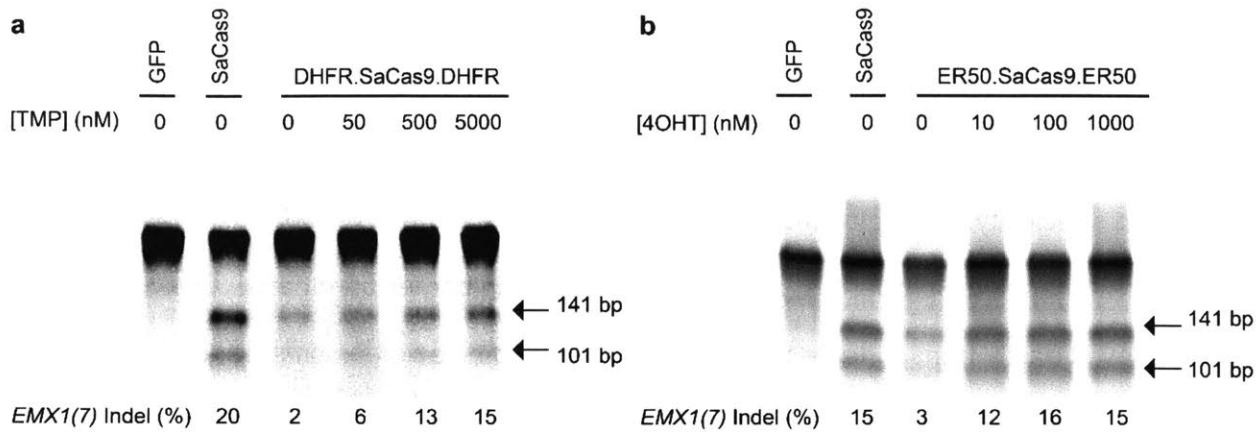

Figure 3.20 | Surveyor assay gels depicting induction of DD-regulated SaCas9 activity by the small molecules TMP or 4OHT

(**a**) Surveyor nuclease assay showing TMP dose-dependent induction of DHFR.SaCas9.DHFR-mediated gene editing activity. HEK293T cells were transfected with SaCas9 or DHFR.SaCas9.DHFR and sgRNA targeted to *EMX1*(7) and treated with the indicated doses of vehicle or TMP for 72 h prior to genomic DNA extraction and analysis by the Surveyor nuclease assay.

(**b**) Surveyor nuclease assay showing dose regulation of ER50.SaCas9.ER50 mediated gene editing activity by the small molecule 4OHT. HEK293T cells were transfected with SaCas9 or ER50.SaCas9.ER50 targeted to *EMX1*(7) and treated with the indicated doses of vehicle or 4OHT for 72 h prior to genomic DNA extraction and analysis by the Surveyor nuclease assay.

3.5 Concluding Remarks

In summary, chemogenic control of Cas9 endowed by DD fusion enables robust control of genome-interrogating activities across multiple dimensions, including dose, time, gene targets, and specificity. The small molecules employed are inexpensive, non-toxic, display favorable pharmacologic properties, and TMP is blood brain barrier-permeable. Our methodology is easily transportable to other cells and to complex organisms[6], enabling not just biomedical applications but also control of CRISPR-gene drives. Finally, our modular approach should prove readily extensible to emerging, next-generation RNA-guided endonucleases.

In the context of studying proteostasis networks and other complex gene networks, we envision that transcriptional interrogation will prove to be an invaluable tool. While in vitro data suggests that stoichiometric balance between components dictates activity of chaperones and other proteostasis network components, it is unclear what role stoichiometric balance alterations of proteostasis network components play in living systems. This lack of clarity is a result of a lack of methods to control the expression levels of components in a robust and scalable manner. Since the DD-regulated dCas9 transcriptional control system is tunable and capable of targeting one or many genes in an orthogonal manner, we can foresee that this method will open doors to the study of proteostasis networks and many other complex biological processes in live cells.

3.6 Methods

3.6.1 Reagents and plasmids

Trimethoprim was purchased from Alfa Aesar (J63053, ≥98% purity, validated by LC-MS and stabilization of a fluorescent positive control construct DHFR.YFP). (*Z*)-4-Hydroxytamoxifen was purchased from Sigma Aldrich (H7904, ≥98% purity, validated by LC-MS and stabilization of a fluorescent positive control construct ER50.YFP). Lipofectamine 2000 (Life Technologies) was used as a transfecting agent according to the manufacturer's protocol. Plasmid sequences for DD-fused Cas9 and transcription activation domain constructs are available upon request.

3.6.2 Cell culture

All cells were cultured at 37 °C in a 5% CO_2 atmosphere. HEK293T cells (Life Technologies) used in transcriptional activation experiments were cultured in Dulbecco's modified Eagle's medium (CellGro) supplemented with 10% fetal bovine serum (CellGro) and 1% penicillin/streptomycin/glutamine (CellGro). HEK293T cells used in surveyor assays and nuclease specificity experiments and U2OS.eGFP-PEST cells[26] stably integrated with an eGFP-PEST fusion gene were maintained in Dulbecco's modified Eagle's medium (Life Technologies) supplemented with 10% FBS, 1x penicillin/streptomycin/glutamax (Life Technologies) and 400 μg/mL of the selection antibiotic G418 (for the U2OS.eGFP-PEST cells). Cells were continuously maintained at <90% confluency.

3.6.3 Transcription activation experiments and quantitative RT-PCR analyses

Transient transfections of HEK293T cells (750,000 cells, 6-well format) were performed with an equivalent mass of each plasmid (dCas9, destabilized domain effector, and sgRNA) for a total mass of 5 μg of plasmid DNA. 24 h post-transfection, cells were treated with the appropriate dose of DD-stabilizing small molecule(s), as indicated, for 18 h prior to harvest and subsequent RNA extraction using the EZNA Total RNA Kit I (Omega). For reversibility experiments, cells were treated 24 h post-transfection with 100 nM TMP for 18 h, at which point cells were treated

with fresh media either containing 100 nM TMP or lacking TMP. Post-media swap, cells were incubated for the indicated times in fresh media prior to harvest and subsequent RNA extraction using the EZNA Total RNA Kit I (Omega). qPCR reactions were performed on cDNA prepared from 1000 ng of total cellular RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). TaqMan qPCR probes (Life Technologies, **Table 3.2**) or other primers (**Table 3.3**) and Fast Advanced Master Mix (Life Technologies) were used in 5 μl multiplexed reactions and 384-well format in a Light Cycler 480 II Real-Time PCR machine. All measurements were performed at least in triplicate. Data were analyzed using the LightCycler® 480 Software, Version 1.5 (Roche) by the $\Delta\Delta C_t$ method: target gene C_t values (FAM dye) were normalized to *GAPDH* C_t values (VIC dye), and fold-changes in target gene expression were normalized to RFP-transfected experimental controls. Analyzed data are reported as the mean ±95% confidence interval.

3.6.4 Next-generation sequencing of Cas9-mediated genome modifications

HEK293T cells (130,000 cells, 24-well plate) were transiently transfected with 400 ng of the indicated Cas9 plasmid and 100 ng of the *EMX1(1)* sgRNA expression plasmid in the presence or absence of TMP or 4OHT, as appropriate[27]. Transfection of an eGFP-encoding plasmid was used as a control. Genomic DNA was extracted 72 h post-transfection using the QuickExtract™ DNA extraction kit (Epicentre) by incubating the cell suspension at 65 °C for 15 min, 68 °C for 15 min, and 98 °C for 10 min. Next-generation sequencing samples were prepared via two-step PCR (see **Table 3.4**) following a previously reported protocol[28]. In the first step, PCR was performed to amplify the target gene of interest and introduce adapters. In the second step, PCR was used to attach Illumina P5 adapters with barcodes, after which PCR products were isolated via gel purification. DNA concentrations were determined using the Qubit® dsDNA HS Assay Kit (Life Technologies) and processed for NGS analysis using the MiSeq Reagent Kit v2 300 (Illumina) according to the manufacturer's protocol.

3.6.5 Analysis of Cas9 nuclease activity via disruption of genomic eGFP-PEST

Approximately 200,000 U2OS.eGFP-PEST cells were nucleofected in duplicate with 500 ng of Cas9 and sgRNA expressing plasmids along with a Td-tomato-encoding plasmid using the SE Cell Line 4D-Nucleofector™ X Kit (Lonza) according to the manufacturer's protocol. Approximately 30,000 transfected cells/well in 5 replicates were plated in a 96-well plate (Corning® 3904 clear bottom) and incubated with the indicated quantities of TMP or 4OHT. for 48 h. Cells were fixed using 4% paraformaldehyde and HCS NuclearMask™ Blue Stain (Life Technologies) was used as the nuclear counter-staining agent. Imaging was performed with an IXM 137204 ImageXpress Automated High Content Microscope (Molecular Devices) at 4× magnification under three excitation channels (blue, green and red) with 9 acquisiton sites per well. Images were analyzed in the MetaXpress software and data were plotted using GraphPad Prism 6.

3.6.6 Western blot analyses

Cells transiently expressing the indicated constructs were incubated either in the absence or presence of stabilizing small molecules (TMP or 4OHT) for 24 h or with the proteasome inhibitor MG-132 for 12 h, prior to harvesting. Cell suspensions were spun down at 1000 × g for 5 min and processed following one of two different protocols: (Protocol 1 – Total Cell Lysis) Cells were resuspended in RIPA buffer (150 mM sodium chloride, 1.0% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, and 50 mM Tris, pH 7.5) and incubated at 4 °C for 10 min. The cell suspensions were then vortexed for 10 min at 4 °C followed by spinning down at 16,000 × g for 15 min at 4 °C. The supernatant was transferred to a fresh tube and processed for immunoblotting. (Protocol 2 – Nuclear Extraction) Cell pellets were resuspended in Buffer A (10 mM HEPES pH 7.5, 50 mM NaCl, 0.5 M sucrose, 0.1 mM EDTA, 0.5% Triton X100, and protease inhibitors) and incubated on ice for 10 min. Cell suspensions were centrifuged at 1000 × g for 5 min at 4 °C and the post-nuclear supernatant was transferred to a separate tube. Cell

pellets were washed once with Buffer A followed by Buffer B (10 mM HEPES pH 7.5, 10 mM KCl, 0.1 mM EDTA, and 0.1 mM EGTA). Finally, the cell pellets were resuspended in Buffer C (10 mM HEPES pH 7.5, 500 mM NaCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.1% IGEPAL (NP40), and protease inhibitors) and vortexed for 15 min at 4 °C. The cell suspensions were centrifuged at 16,000 × g at 4 °C. The resulting nuclear extract in the supernatant was transferred to a fresh tube and processed for immunoblotting. In a typical immunoblotting protocol, 40 μg of normalized proteins were resuspended in the appropriate lysis buffer and electrophoresed on an 8% Bis/Tris gel with SDS-Tris running buffer. The protein bands were transferred to a nitrocellulose membrane and probed with α-HSF1(c) (Abcam; ab52757; 1:100,000), α-Cas9 (Abcam; ab191468; 1:1,000), α-actin as a loading control for total cell lysates (Sigma; a1978; 1:10,000), and/or α-H3K27me2 as a loading control for nuclear extracts (Cell Signaling Technologies; 9728; 1:1,000). Following blocking and incubation with primary antibodies, membranes were incubated with 680 or 800 nm fluorophore-labeled secondary antibodies (Li-COR Biosciences; 1:10,000) prior to detection using a LI-COR Biosciences Odyssey Imager. Images were processed in Image Studio Lite version 3.1.4 from LI-COR Biosciences.

3.6.7 Surveyor nuclease assay

The small molecule-mediated control of the DHFR.SaCas9.DHFR and ER50.SaCas9.ER50 gene editing systems were investigated by the SURVEYOR nuclease assay[29]. Briefly, HEK293T cells (130,000 cells/well in a 24-well plate) were transiently transfected with 150 ng of either SaCas9 or DD.SaCas9.DD constructs (DHFR.SaCas9.DHFR or ER50.SaCas9.ER50) along with an *EXM1(7)*-targeting sgRNA-expressing plasmid. The cells were incubated with or without appropriate DD-stabilizing small molecules (TMP: 0, 50, 500, or 5000 nM; 4OHT: 0, 10, 100, or 1000 nM) for 72 h post-transfection at 37 °C. Following cell harvesting, the genomic DNA was isolated using the QuickExtract™ DNA extraction kit (Epicentre). Genomic DNA was then subjected to PCR using primers corresponding to 157 bp in the *EMX1(7)* gene segment (**Table**

3.5) and the amplicons were purified by the QIAQuick PCR purification kit. The isolated amplicons were normalized and subjected to a quick-annealing protocol (ramp 0.03 °C/s), following which they were incubated with Surveyor nuclease S (Surveyor Mutation Detection Kits, IDT) at 42 °C for 1 h. The samples were analyzed by running on a TBE gel and the cleavage bands were visualized by staining with SYBR Gold. Indel frequencies were calculated using the equation "indel frequency = 100*(1-sqrt(1-((b+c)/(a+b+c))))", where a is the intensity of undigested PCR band and b and c are the intensities of the cleaved bands.

Table 3.2 | TaqMan® probes used for qPCR assays

Gene	ID
ASCL1	Hs00269932_m1
IL1RN	Hs00893626_m1

Table 3.3 | Primers used for SYBR Universal Master Mix qPCR assays

Gene	Forward Primer	Reverse Primer
RPLP2 (housekeeping)	CGTCGCCTCCTACCTGCT	CCATTCAGCTCACTGATAACCTTG
NANOG	GATTTGTGGGCCTGAAGAAA	CAGATCCATGGAGGAAGGAA

Table 3.4 | Primers used to generate amplicons for next-generation sequencing

Gene	Name	5'-Sequence-3'		Forward Primers	Reverse Primers
		Target	PAM		
EMX1	EMX1(1)	GAGTCCGAGCAGAAGAAGAA	GGG	CCATCTCATCCCTGCGTGTCTCc CAAAGTACAAACGGCAGAAGC	CCATCTCATCCCTGCGTGTCTCc GTTGCCCACCCTAGTCATTG
	OT1	GAGTTAGAGCAGAAGAAGAA	AGG	CCATCTCATCCCTGCGTGTCTCc TTCTGAGGGCTGCTACCTGT	CCATCTCATCCCTGCGTGTCTCc GCCCAATCATTGATGCTTTT
	OT2	GAGTCTAAGCAGAAGAAGAA	GAG	CCATCTCATCCCTGCGTGTCTCc CACGGCCTTTGCAAATAGAG	CCATCTCATCCCTGCGTGTCTCc GGCTTTCACAAGGATGCAGT
	OT3	GAGTCCTAGCAGGAGAAGAA	GAG	CCATCTCATCCCTGCGTGTCTCc CCAGACTCAGTAAAGCCTGGA	CCATCTCATCCCTGCGTGTCTCc TGGCCCCAGTCTCTCTTCTA
	OT4	AAGTCTGAGCACAAGAAGAA	TGG	CCATCTCATCCCTGCGTGTCTCc GTTCTGACATTCCTCCTGAGGG A	CCATCTCATCCCTGCGTGTCTCc ATGGCTTACATATTTATTAGATAAAATGTATTCC
	OT5	GAGGCCGAGCAGAAGAAAGA	CGG	CCATCTCATCCCTGCGTGTCTCc TGGGAGAGAGACCCCTTCTT	CCATCTCATCCCTGCGTGTCTCc TCCTGCTCTCACTTAGACTTTCTC
VEGFA	VEGFA(1)	GGTGAGTGAGTGTGTGCGTG	TGG	CCATCTCATCCCTGCGTGTCTCc GCGTCTTCGAGAGTGAGGAC	CCATCTCATCCCTGCGTGTCTCc GGGGAGAGGGACACACAGAT
	OT1	GCTGAGTGAGTGTATGCGTG	TGG	CCATCTCATCCCTGCGTGTCTCc GCCCATTTCTCCTTTGAGGT	CCATCTCATCCCTGCGTGTCTCc AGCCACAGAGGTGGAGACTG

Table 3.5 | Primers used to generate amplicons for Surveyor nuclease assay

Gene	5'-Sequence-3'		Forward Primer	Reverse Primer
	Target	PAM		
EMX1(7)	GGCCTCCCCAA AGCCTGGCCA	GGGAGT	CCATCTCATCCCTGCGTGTCT CCAACCCACGAGGGCAGAGT	CCTCTCTATGGGCAGTCGGTGAT GGAGGAGAAGGCCAAGTGGTC

3.7 References

1. Doudna, J.A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
2. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-78 (2014).
3. Gantz, V.M. & Bier, E. The dawn of active genetics. *Bioessays* **38**, 50-63 (2016).
4. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-91 (2013).
5. Hilton, I.B. et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33**, 510-7 (2015).
6. Dominguez, A.A., Lim, W.A. & Qi, L.S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5-15 (2016).
7. Nunez, J.K., Harrington, L.B. & Doudna, J.A. Chemical and biophysical modulation of Cas9 for tunable genome engineering. *ACS Chem. Biol.* ***11***, 681-688 (2016).
8. Nguyen, D.P. et al. Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat. Commun.* **7**, 12009 (2016).
9. Oakes, B.L. et al. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* 646-651 (2016).
10. Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D. & Wandless, T.J. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **17**, 981-8 (2010).
11. Miyazaki, Y., Imoto, H., Chen, L.C. & Wandless, T.J. Destabilizing domains derived from the human estrogen receptor. *J. Am. Chem. Soc.* **134**, 3942-5 (2012).
12. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G. & Wandless, T.J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004 (2006).
13. Moore, C.L. et al. Transportable, chemical genetic methodology for the small molecule-mediated inhibition of heat shock factor 1. *ACS Chem. Biol.* **11**, 200-10 (2016).
14. Shoulders, M.D., Ryno, L.M., Cooley, C.B., Kelly, J.W. & Wiseman, R.L. Broadly applicable methodology for the rapid and dosable small molecule-mediated regulation of transcription factors in human cells. *J. Am. Chem. Soc.* **135**, 8129-32 (2013).
15. Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973-6 (2013).
16. Balboa, D. et al. Conditionally stabilized dCas9 activator for controlling gene expression in human cell reprogramming and differentiation. *Stem Cell Rep.* **5**, 448-59 (2015).
17. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-8 (2015).
18. Zalatan, J.G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-50 (2015).
19. Zetsche, B., Volz, S.E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139-142 (2015).
20. Nishimasu, H. et al. Crystal structure of Staphylococcus aureus Cas9. *Cell* **162**, 1113-26 (2015).

21. Esvelt, K.M., Smidler, A.L., Catteruccia, F. & Church, G.M. Concerning RNA-guided gene drives for the alteration of wild populations. *eLife* **3**, e03401 (2014).
22. Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A. & Liu, D.R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316-8 (2015).
23. Tsai, S.Q. & Joung, J.K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.* **17**, 300-312 (2016).
24. Zuris, J.A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73-80 (2015).
25. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012-1019 (2014).
26. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822-6 (2013).
27. Schneeberger, K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* **15**, 662-76 (2014).
28. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827-32 (2013).
29. Ran, F.A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281-308 (2013).

Chapter 4: A Processive Protein Chimera Introduces Mutations Across Defined DNA Regions *in vivo*

This chapter is adapted from the following manuscript:
Moore, C.L.,[1] Papa III, L.J.,[1] Shoulders, M.D. *in submission*
[1]These authors contributed equally to this work and are listed in alphabetical order.

4.1 Author Contributions

C.L.M., L.J.P., and M.D.S. planned research, designed experiments, and analyzed data. L.J.P. performed molecular cloning and seamless recombineering to generate vectors and vectors and bacterial strains with assistance from C.L.M. L.J.P. optimized mutagenesis experiment conditions with assistance from C.L.M. C.L.M. and L.J.P. performed drug resistance assays. C.L.M. maintained continuous culture samples and prepared samples for sequencing with assistance from L.J.P. The MIT BioMicro Center performed Illumina sequencing. We thank Vincent Butty for downstream analysis of sequencing data. M.D.S. supervised research.

4.2 Abstract

Laboratory-timescale evolution relies on the generation of large, mutationally diverse libraries to rapidly explore biomolecule sequence landscapes. Although numerous *in vitro* mutagenesis techniques are available, *in vivo* mutagenesis is limited[1]. Global mutagenesis methods are capable of increasing mutation rates *in vivo*, but unfortunately introduce extensive lethal mutations that constrain library size[2]. Moreover, when off-target mutations appear in genes that allow the organism to circumvent or "cheat" a selection, DNA libraries become contaminated with problematic, uninteresting false positives[2-4]. Here, we report the development and application of the MutaT7 chimera, a potent and highly targeted mutagenesis system. The MutaT7 system utilizes a DNA-damaging cytidine deaminase fused to a processive RNA polymerase to continuously direct mutations to specific, well-defined DNA regions of any relevant length. MutaT7 thus provides, for the first time, a mechanism for *in vivo* targeted mutagenesis across multi-kb DNA sequences. MutaT7 can be utilized in a broad range of organisms, opening the door to new types of evolution experiments.

4.3 Introduction to Targeted Mutagenesis *in vivo*

Traditional *in vivo* mutagenesis strategies, which are especially important for studying and using evolution in living systems, rely on exposing organisms to exogenous mutagens (e.g., high energy light or chemicals[5, 6]) or expressing mutagenic enzymes (e.g., XL1-Red[7] or the MP6 plasmid[2]). These global mutagenesis strategies can yield high mutation rates and diverse genetic landscapes. However, the extensive occurrence of mutations throughout the genome is problematic in many contexts, especially in directed evolution experiments. For example, off-target mutations outside the intended DNA region are often toxic when they occur in the many essential portions of the genome (**Figure 4.2**) [8, 9], a problem that can severely limit library size and lead to rapid silencing of mutagenic plasmids. Moreover, global mutagens also introduce "parasite" variants into DNA libraries, which originate from undesired mutations outside the gene of interest that allow an organism to circumvent selection schemes (**Figure 4.3**)[4]. More refined implementations of chemical mutagens have illustrated the experimental benefits of limiting off-target mutations, but such methods are laborious, require extensive engineering, and present significant health hazards[10].

Targeted *in vivo* mutagenesis strategies have the potential to overcome these deficiencies. For example, DNA-damaging enzymes fused to deactivated Cas9 nucleases can edit bases at specific genetic loci while minimizing off-target mutations[11-14]. Cas9-based approaches hold potential, but unfortunately require significant engineering to tile mutagenic enzymes throughout a target DNA that may be several kb in length or even longer[15, 16]. Moreover, this engineering likely needs to be repeated after each successive round of evolution introduces new mutations in the target DNA.

We rationalized that a processive DNA-traversing biomolecule tethered to a DNA-editing enzyme could provide a general solution to the problem of directing mutations across defined DNA regions. Monomeric RNA polymerases possess inherently high promoter specificity[17] and high processivity during transcription[18]. Cytidine deaminases are potent DNA-damaging

enzymes that can act on ssDNA substrates during transcription[18, 19]. We envisioned that merging the unique features of these two enzyme classes by creating a chimeric "MutaT7" protein consisting of a cytidine deaminase (rApo1) fused to T7 RNA polymerase (T7-pol) would allow us to target mutations to any DNA region lying downstream of a T7 promoter (**Figure 4.4**).

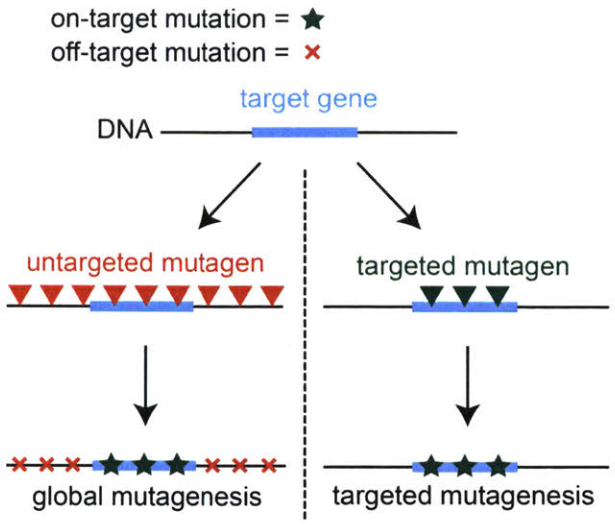

Figure 4.1 | Comparison of global and targeted mutagenesis

Schematic highlighting differences between global versus targeted mutagenesis and the associated outcomes

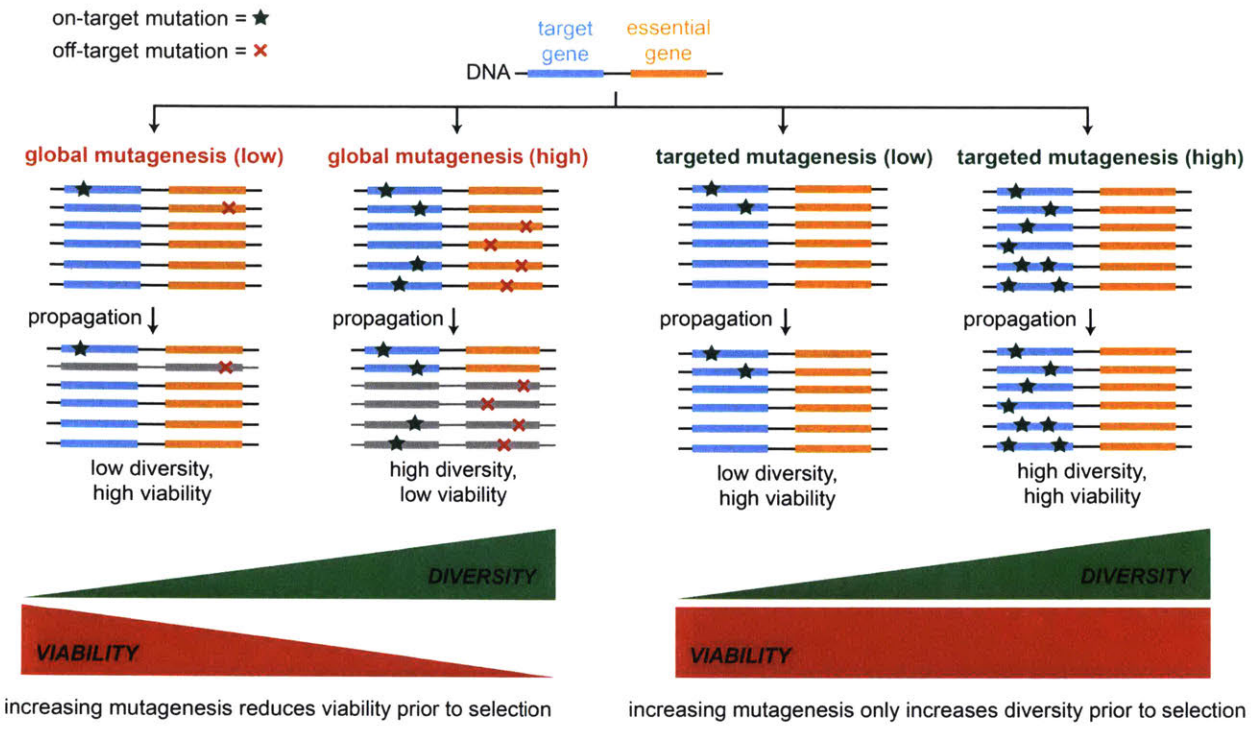

Figure 4.2 | Global mutagenesis causes deleterious mutations in essential genes

Diagram illustrating how global mutagenesis reduces viability by indiscriminately introducing mutations across an organism's entire genome. As global mutagens incorporate mutations in target genes and other genes such as essential genes, attempts to increase the global mutagenesis rate and thus library diversity lead to decreased cell viability. Targeted mutagenesis allows for a high mutagenesis rate that does not decrease cell viability by minimizing off-target mutations in essential genes.

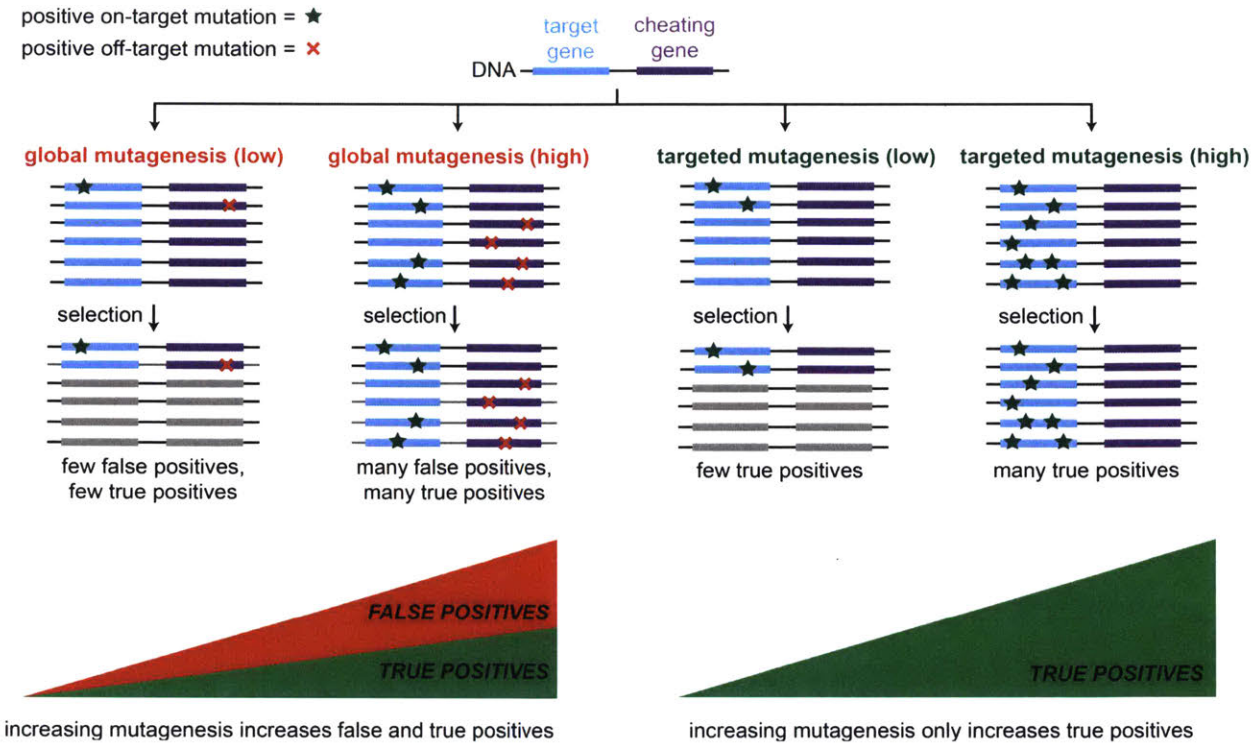

Figure 4.3 | Off-target mutations promotes cheating of selection

Schematic showing how global mutagenesis incorporates off-target mutations in genes that allow an organism to cheat a phenotypic selection. In an evolution experiment, this will cause a certain rate of false positives. Targeted mutagenesis minimizes these false positives by preventing off-target mutations in genes that allow the organism to cheat the selection.

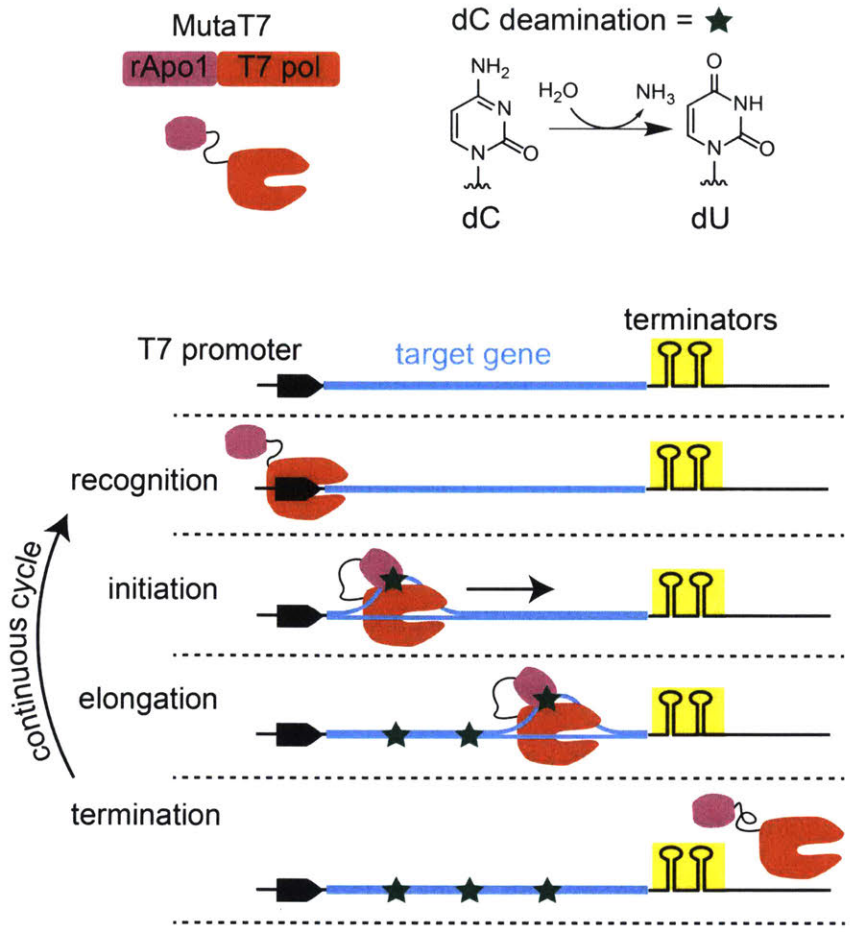

Figure 4.4 | MutaT7: an approach that incorporates mutations throughout target DNA

Diagram (above) of the MutaT7 construct, which consists of a cytidine deaminase fused to T7 polymerase. A schematic (below) demonstrates the processive cycle through which targeted mutagenesis is achieved in a target region of DNA.

4.4 Results

4.4.1 MutaT7 Mutagenizes DNA Between the T7 Promoter and a Termination Element

To begin, we fused rApo1 to the N-terminus of T7-pol, because the carboxy group of the T7-pol C-terminus is implicated in catalysis during the elongation phase[20]. To maintain a stable and low expression level, we used seamless recombineering to integrate the MutaT7 gene under control of a weak promoter into the genome of *E. coli* lacking uracil DNA glycosylase (*Δung*) (**Figure 4.5**, **Supplementary Table 4.1**). We next assayed for targeted mutagenesis using a codon reversion assay based on bacteria artificial chromosome reporter plasmids either having or lacking a T7 promoter sequence upstream of silent drug resistance genes with ACG triplets in place of their ATG start codons (**Figures 4.6 – 4.10**). The kanamycin resistance gene *Kan*^{R} was placed immediately downstream of the T7 promoter. In this assay, successful C→T mutagenesis at the start codon yields kanamycin-resistant colonies. Global mutagens such as the MP6 plasmid yielded high levels of Kan-resistant colonies regardless of the presence or absence of a T7 promoter, consistent with a lack of targeting (**Figure 4.10**, orange solid bars). In contrast, in MutaT7 strains significant kanamycin resistance was only observed using reporter plasmids having a T7 promoter upstream of the *Kan*^{R} gene, a result consistent with successful targeted mutagenesis. Importantly, expression of a catalytically dead version of MutaT7 (drApo1–T7) lacking a critical residue for cytidine deaminase activity[21] yielded kanamycin resistance frequencies similar to background levels, indicating that T7 activity was not responsible for increased kanamycin resistance (**Figure 4.10**, red solid bars).

T7 promoter-dependent *Kan*^{R} mutagenesis by MutaT7 shows that mutagenesis can be targeted to a desired DNA region near a T7 promoter. Because T7 polymerase is highly processive, we anticipated mutations would also be introduced further downstream of the T7 promoter. The presence in our reporter plasmid of a tetracycline-resistance (*Tet*^{R}) gene with an inactive, ACG start codon separated from the *Kan*^{R} gene by an ~1.6 kbp spacer DNA sequence (**Figure 4.6**) provided a mechanism to assay for such processivity. We observed high levels of

MutaT7-dependent tetracycline resistance only in reporter strains having the T7 promoter, consistent with targeted and processive introduction of mutations across a lengthy, multi-kb DNA region (**Figure 4.10**, red striped bars). Once again, global mutagens generated tetracycline-resistant colonies at a high frequency in all reporter plasmids (**Figure 4.10**, orange striped bars).

Targeted mutagenesis using the processive MutaT7 chimera requires not just recruitment to a DNA locus, but also termination upon reaching the end of the DNA region of interest. To address termination, we used Kan^R/Tet^R reporter plasmids in which we replaced the DNA spacer with one or more T7 terminators, and then assayed for both kanamycin and tetracycline resistance (**Figure 4.6**). We found that four copies of the T7 terminator were sufficient to prevent mutagenesis beyond the DNA of interest (**Figure 4.11**). Using the T7 terminator array, we observed tetracycline resistance for MutaT7 strains similar to background levels, while kanamycin resistance remained high (**Figure 4.10** and **Figure 4.11**). Global mutagens again induced high levels of kanamycin- and tetracycline-resistance, regardless of the presence of a T7 terminator array (**Figure 4.10**; see also **Supplementary Tables 4.2a–c**).

Figure 4.5 | Promoter design to reduce expression of mutaT7

a) The $P_{A1lacO-1}$ promoter has been engineered to have minimal leaky expression when repressed with lacI[22]. The lac operators are shown in blue and the σ^{70} binding sites are shown in red. The BBa_J23114 promoter from the Anderson Collection [23] has been shown to have about a tenth of the strength of the σ^{70} consensus binding sites. With the intention of obtaining a weak, strongly repressed promoter, the σ^{70} binding sites of BBa_ J23114 were cloned onto $P_{A1lacO-1}$ to yield $P_{A1lacO-Tenth}$ (changes shown in green).

b) In order to increase the expression of lacI from the DH10B genome, the endogenous P_{lacI} promoter was replaced with the strong, constitutive P_{tac} promoter[24]. The σ^{70} binding sites are shown in red.

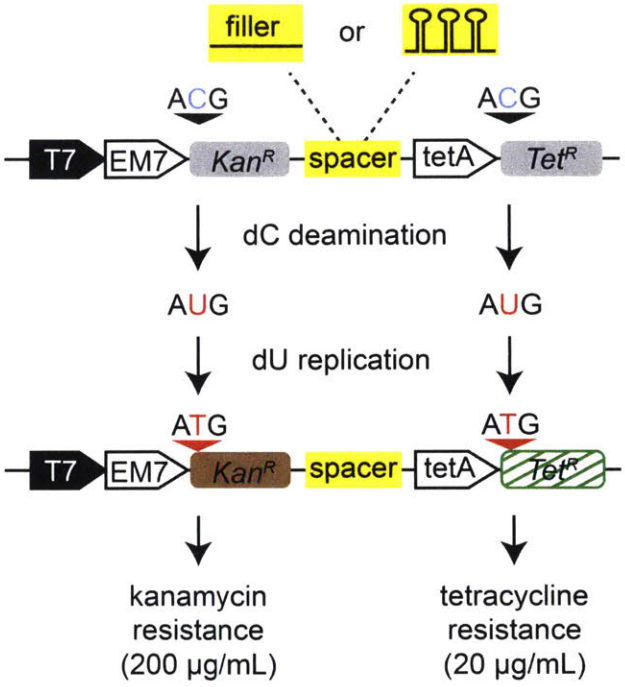

Figure 4.6 | Design of reporter constructs in start codon reversion drug resistance assay

Schematic of a drug resistance start codon reversion reporter assay for measuring extent of mutagenesis at specific DNA loci. The first gene (*Kan*^{R}) reports on-target mutagenesis, while the second gene (*Tet*^{R}) reports off-target mutagenesis downstream of a stretch of filler DNA or a terminator array that halts MutaT7.

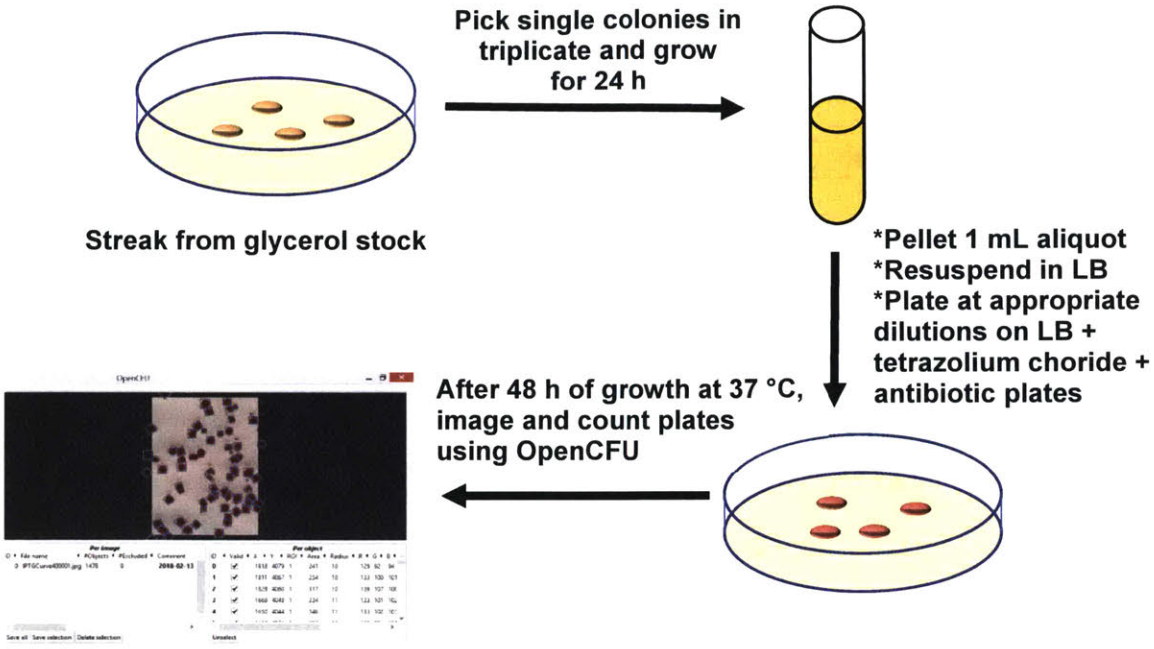

Figure 4.7 | Drug resistance mutation assay workflow

The mutation assay workflow is shown. Glycerol stocks of each sample were streaked on LB agar with appropriate antibiotics and grown at 37 °C for 24 h to obtain clones. Single colonies were picked in triplicate and grown in LB with appropriate antibiotics and inducers of mutagenesis at 37 °C for 24 h to accumulate mutations. 1 mL aliquots of each culture were pelleted and resuspended in LB to remove antibiotics and inducers. The resuspension was plated at various dilutions on plates with various antibiotics to analyze the mutation rates and cell viability. The plates also contained a metabolic dye, tetrazolium chloride, for contrast during imaging. After incubating at 37 °C for 48 h, the plates were imaged on a document scanner at 400 dots per inch and colonies were counted using the OpenCFU (3.9.0) software[25].

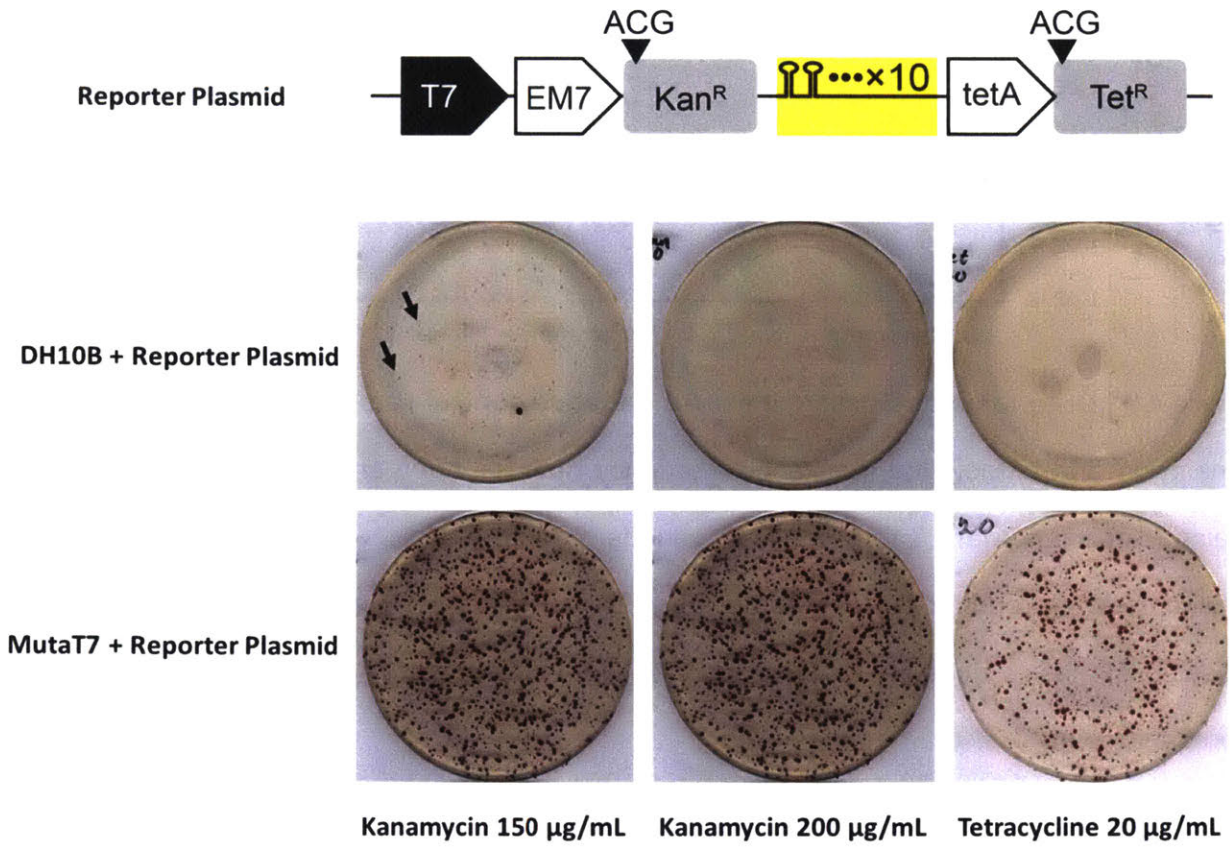

Figure 4.8 | Optimizing antibiotic concentrations for mutation assays

At concentrations of less than 200 μg/mL, small colonies (black arrows) appeared on LB + kanamycin + tetrazolium chloride plates with DH10B carrying the reporter plasmid. The small colonies theoretically may have been present owing to a very low level of expression of the kanamycin resistance gene through translation initiation from the ACG start codon[26]. On plates with 200 μg/mL kanamycin, the small colonies on the DH10B plate did not appear even after 48 h. The number of colonies on plates of MutaT7 cells (**Supplementary Table 4.1**) with the reporter plasmid were similar between plates with 150 μg/mL and 200 μg/mL kanamycin. At a concentration of 20 μg/mL tetracycline, no colonies appeared on LB + tetracycline + tetrazolium chloride plates with DH10B carrying the reporter plasmid, while many colonies appeared with MutaT7 carrying the reporter plasmid.

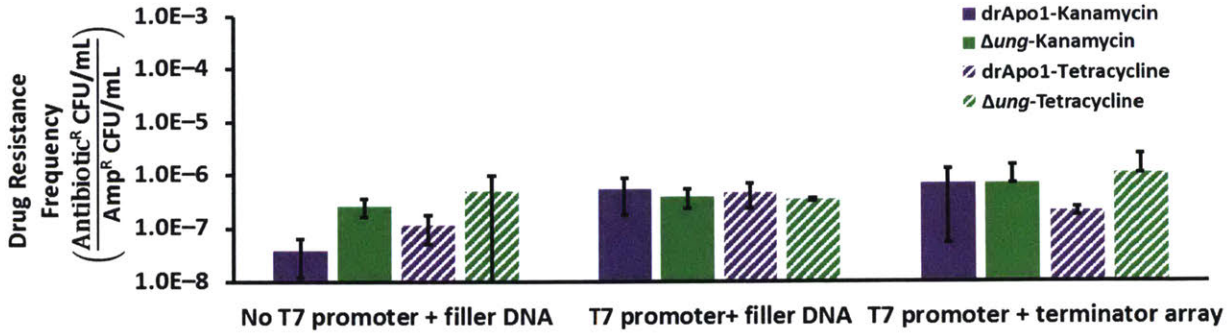

Figure 4.9 | Drug resistance assay data for negative control samples

Kanamycin and tetracycline resistance frequency data for Δ*ung* and drApo1 negative control strains (**Supplementary Table 4.1**) with various reporter plasmids suggest that neither strain mutagenizes the reporter plasmid appreciably. Experiment performed as in **Figure 1d** with the indicated strains. Values represent mean of independent experiments (n = 3); error bars represent s.e.m.

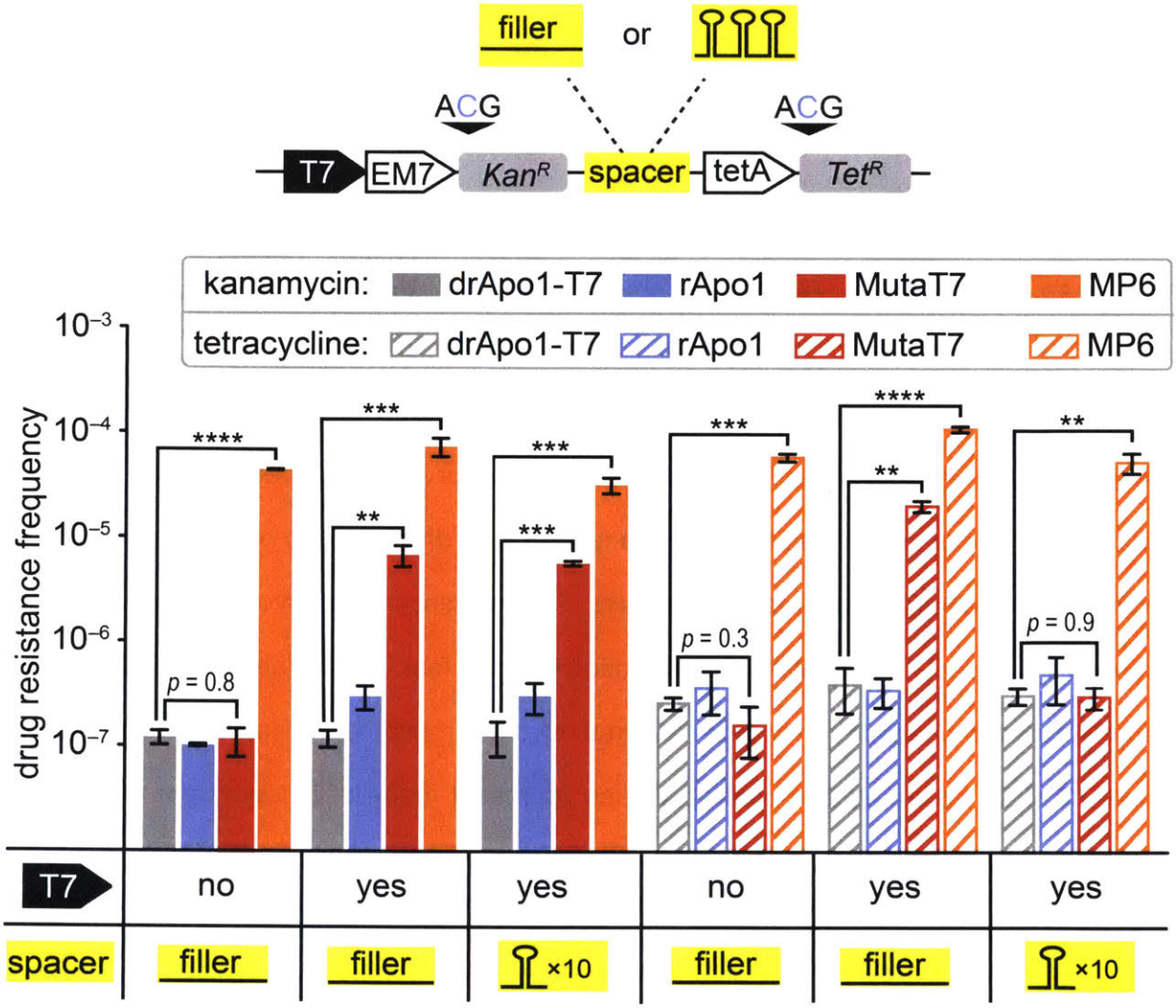

Figure 4.10 | Codon reversion assays demonstrate targeted mutagenesis for mutaT7

Codon reversion reporter assay data for different combinations of mutagen and reporter plasmids. Mutagens include deactivated rApo1 fused to T7 RNA polymerase (drApo1–T7, negative control), unfused rApo1 (rApo1), targeted mutagen (MutaT7), and global mutagen (MP6). Kanamycin and tetracycline resistance frequencies are represented by solid and striped bars, respectively. Values represent mean of independent experiments (n = 3); error bars represent s.e.m.; statistical significance was evaluated by a Student's t-test; *p < 0.05, **p < 0.01 and ***p < 0.001; notable non-significant p-values also shown.

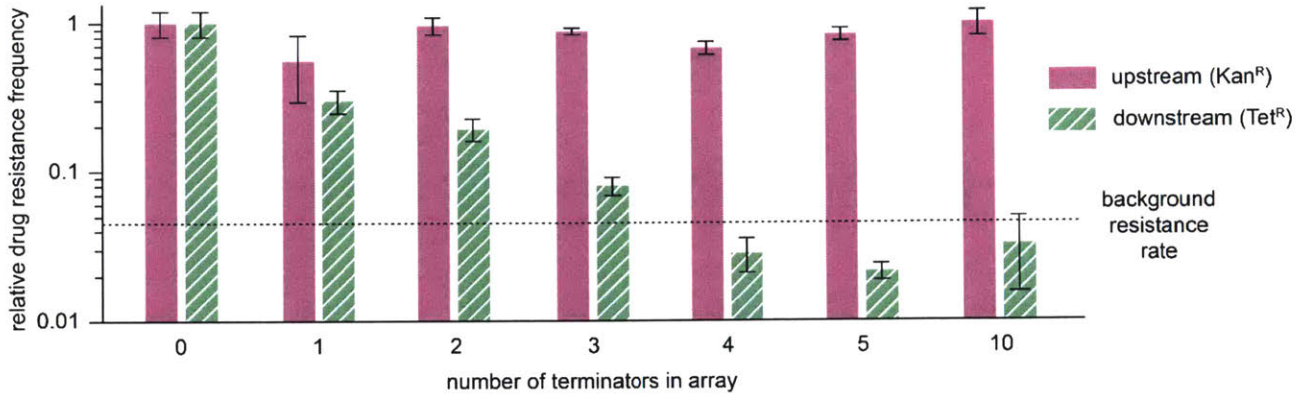

Figure 4.11 | Multiple T7 terminators prevent downstream mutations

After growing the reporter plasmid in the MutaT7 strain for 24 h, the frequency of kanamycin resistant mutant colonies was relatively constant regardless of the number of terminators between the kanamycin and tetracycline resistance genes. The frequency of tetracycline resistant colonies decreased as more T7 terminators were introduced. After four T7 terminators were added, the tetracycline resistance frequency was restored to background levels (as evaluated using a drApo1–T7 strain as a negative control. Values represent mean of independent experiments (n = 3); error bars represent s.e.m.;

4.4.2 Off-target Mutagenesis of Genomic DNA is Minimal with MutaT7

To further assess whether our MutaT7 chimera induces mutagenesis specifically on the target DNA region of interest, we evaluated the emergence of bacterial colonies resistant to rifampicin[27] and fosfomycin[28]. Because resistance to these two drugs can emerge by diverse genomic mutations, the appearance of resistant colonies correlates with off-target mutation rates in the genome[2, 27], and these changes are analogous to the emergence of cheating parasites in directed evolution schemes. Growth of *E. coli* on either rifampicin- or fosfomycin-treated plates revealed that MutaT7-expressing samples displayed drug resistance frequencies comparable to background levels, as opposed to the high frequencies of antibiotic resistance which appeared in all global mutagenesis samples (**Figures 4.12–4.14**).

An important theoretical advantage of targeted mutagenesis is the ability to attain much larger viable library sizes by avoiding off-target, toxic mutations in essential genes outside the DNA region of interest. Based on the apparently low off-target mutagenesis rate of MutaT7, we hypothesized that *E. coli* carrying MutaT7 would have significantly higher viability than bacteria treated with global mutagens. Indeed, consistent with prior work[2], we observed very low viability in all populations treated with global mutagens. In contrast, populations expressing MutaT7 possessed viability similar to untreated cells (**Figures 4.15 and 4.16**). We also found that the total number of kanamycin-resistant colonies was similar between MutaT7 and globally mutagenized samples (**Figure 4.17**) despite the relatively lower mutagenesis rate of the mutaT7 construct compared to MP6 (previously, **Figure 4.6**), highlighting the beneficial effect that minimizing off-target mutations has on library size for *in vivo* evolution schemes.

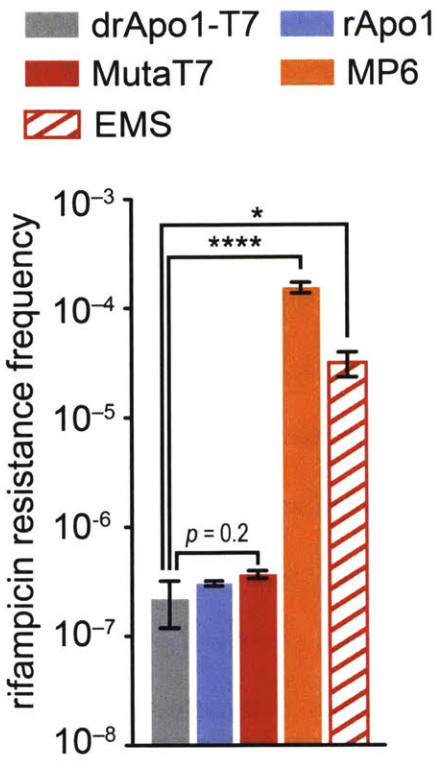

Figure 4.12 | MutaT7 confers few off-target mutations, resulting in low rifampicin resistance frequency

Extent of off-target mutagenesis assessed by rifampicin resistance assay for populations used in **Figure 4.10** carrying the reporter plasmid with terminator array. Data are shown for different mutagenic constructs (solid bars) or chemical mutagen ethyl methanesulfonate (EMS) treatment (candy-stripe bar). Values represent mean of independent experiments (n = 3); error bars represent s.e.m.; statistical significance was evaluated by a Student's t-test; *p < 0.05, **p < 0.01 and ***p < 0.001; notable non-significant p-values also shown.

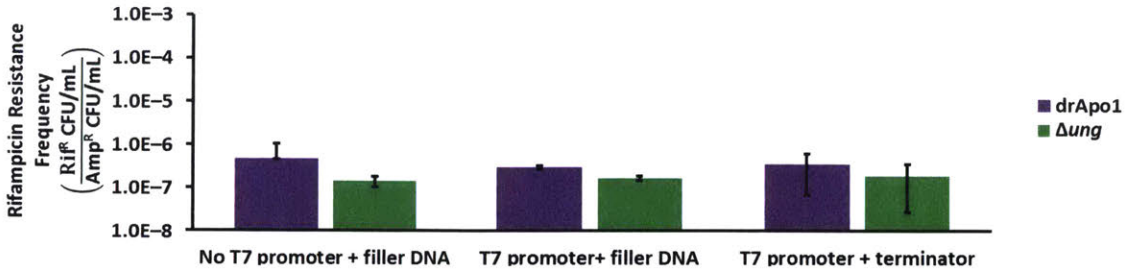

Figure 4.13 | Rifampicin resistance frequency is minimal for negative control samples

Rifampicin resistance frequency data for Δ*ung* and drApo1 negative control strains with various reporter plasmids suggest that neither strain mutagenizes the *E. coli* genome appreciably. Experiment performed as in **Figure 1e** with the indicated strains. Values represent mean of independent experiments (n = 3); error bars represent s.e.m.

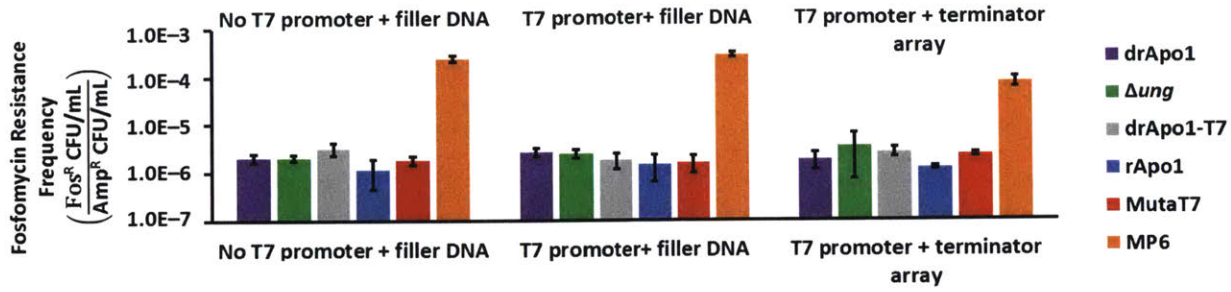

Figure 4.14 | MutaT7 confers few off-target mutations, resulting in low resistance frequency to fosfomycin treatment.

Fosfomycin resistance frequency data show a high mutagenesis rate only in the presence of MP6, suggesting that neither MutaT7 nor the negative controls mutagenize the *E. coli* genome appreciably. Experiment performed as in **Figure 4.12** except cells were plated on LB agar with 100 μg/mL fosfomycin and 50 μg/mL tetrazolium chloride. Values represent mean of independent experiments (n = 3); error bars represent s.e.m.

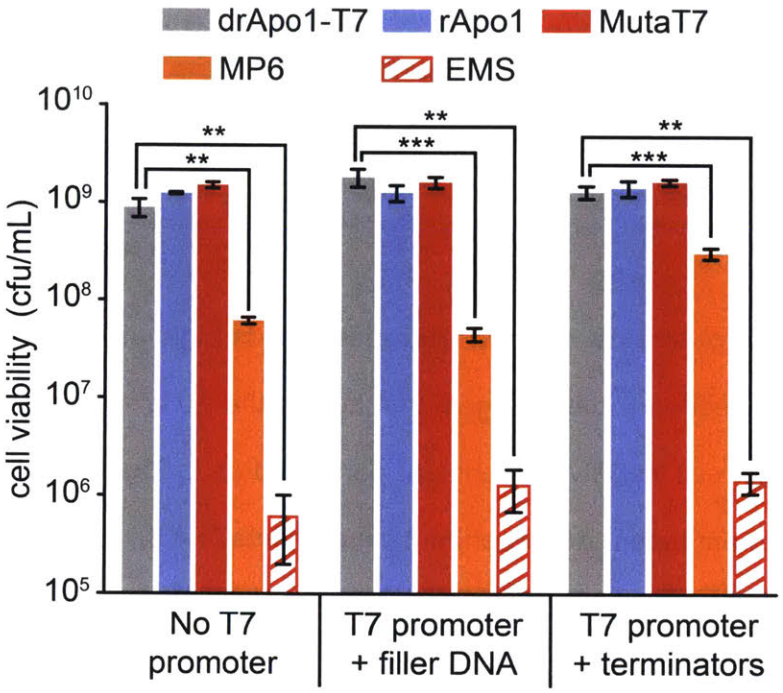

Figure 4.15 | Off-target mutations from global mutagens reduce viability

Viability data for populations of cells in **Figure 4.10** for different mutagenic constructs (solid bars) or EMS treatment (striped bar). Values represent mean of independent experiments (n = 3); error bars represent s.e.m.; statistical significance was evaluated by a Student's t-test; *p < 0.05, **p < 0.01 and ***p < 0.001; notable non-significant p-values also shown.

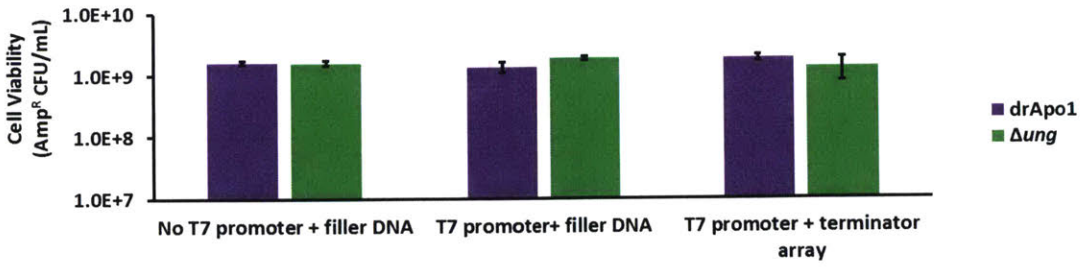

Figure 4.16 | Viability of strains with reporter plasmids is high without any treatment

Ampicillin resistance frequency data suggest that neither the Δ*ung* nor drApo1 negative control strains suffer from low cell viability. Experiment performed as in **Figure 4.15** with the indicated strains. Values represent mean of independent experiments (n = 3); error bars represent s.e.m.

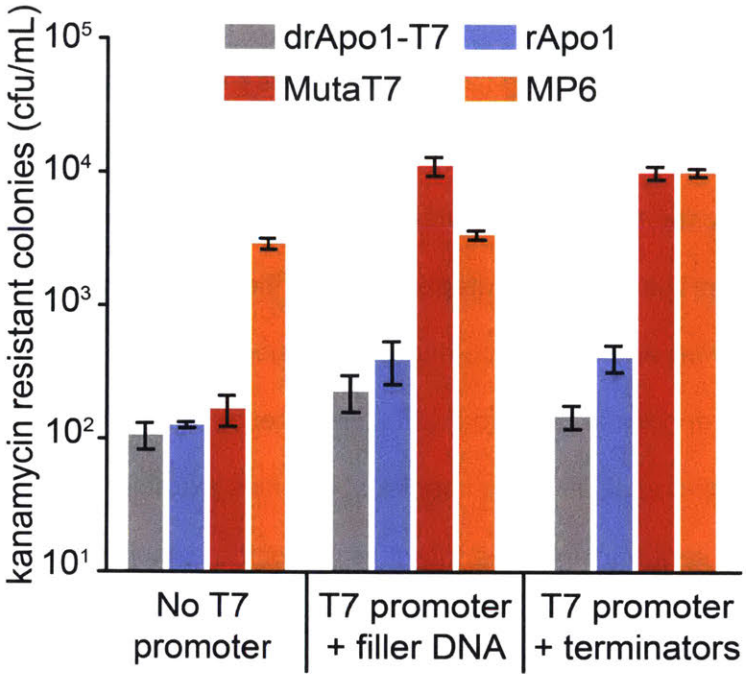

Figure 4.17 | Library size from MutaT7 mutagenesis is comparable to global mutagenesis

Total number of kanamycin resistant colonies for populations in **Figure 1d**, illustrating the effective library size for each sample. Values represent mean of independent experiments (n = 3); error bars represent s.e.m.

4.4.3 Sequencing Reveals MutaT7 Increases Mutation Rates at Multiple Loci in Target DNA

Start codon reversion and drug resistance assays revealed that MutaT7 targets mutations specifically to genes downstream of a T7 promoter, and that the scope of mutagenesis can be constrained by a terminator array. We next turned to DNA sequencing to better understand the processivity of mutagenesis and the extent of on-target versus off-target mutagenesis. We allowed an *E. coli* population expressing MutaT7 and our episomally expressed Kan^R/Tet^R reporter plasmid to drift in the absence of selection pressure for 15 days, and then we isolated episomal DNA from bacterial colonies extracted from the population (**Figure 4.18**). Sanger sequencing of the target region of the episomal DNA revealed that mutations appeared at multiple sites across the entire span of the Kan^R target gene, independent of selection pressure (**Figure 4.19**). We further tested for processive mutagenic activity in a separate experiment by sequencing streptomycin-resistant variants of an initially sensitive allele of *rpsL* downstream of a T7 promoter on a reporter plasmid isolated from a MutaT7-expressing strain of *E. coli* (**Figure 4.20**). Sanger sequencing of isolates again revealed that multiple mutations appeared throughout the targeted *rpsL* gene following propagation in the MutaT7 strain (**Figure 4.21**).

Next, we employed Illumina sequencing to identify mutations anywhere in the episomal reporter DNA sequence obtained from 36 clones drawn from the same *E. coli* population used in **Figure 4.18**. This experiment assesses on- versus off-target mutagenesis across a ~10 kb stretch of DNA containing only ~1 kb of intended target DNA. We used the same strategy to assess mutagenesis in a control *E. coli* population not treated with any mutagen, and a population subjected to global mutagenesis. We observed that clones drawn from MutaT7 samples displayed many mutations throughout the episome when the terminator array was removed but the T7 promoter was maintained (**Figure 4.22**). Treatment with the MP6 global mutagen also led to mutations across the entire episome. In contrast, mutations in MutaT7 samples appeared almost exclusively within the Kan^R target gene when both a promoter and

terminator array were present, even after 15 days of continuous culturing (**Figure 4.18**). Upon normalizing on- and off-target mutation rates, we observed that the few off-target mutations found on plasmids with a terminator from MutaT7 strains were present only to the same extent as in the control sample not treated with any mutagen (**Figure 4.23**, red striped bars). Quantitation of the on-target to off-target mutation ratios confirmed that MutaT7 produced by far the highest ratio of on-target mutations (**Figure 4.24**).

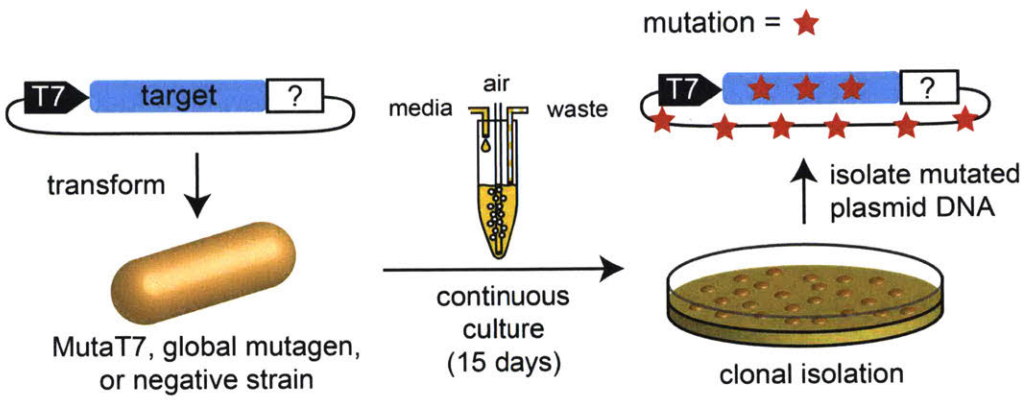

Figure 4.18 | Experimental design to measure mutations throughout episomal DNA

Diagram of reporter construct and continuous culture experiment to assess mutation accumulation under drift conditions. Reporter episomal DNA is carried during continuous culture for 15 days without selection pressure. After cloning, episomal DNA is isolated for further assessment of mutations.

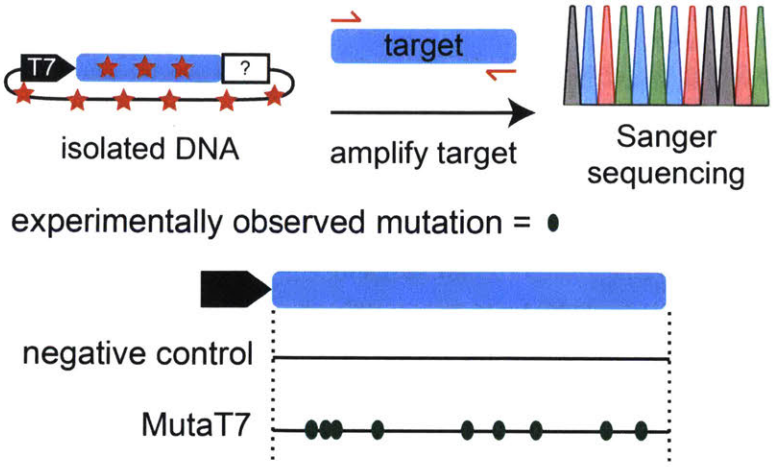

Figure 4.19 | MutaT7 mutates target DNA at multiple loci during continuous culturing

Schematic and representation of mutations observed by Sanger sequencing 96 clones in the indicated cell populations following 15 d of continuous growth in the absence of selection pressure.

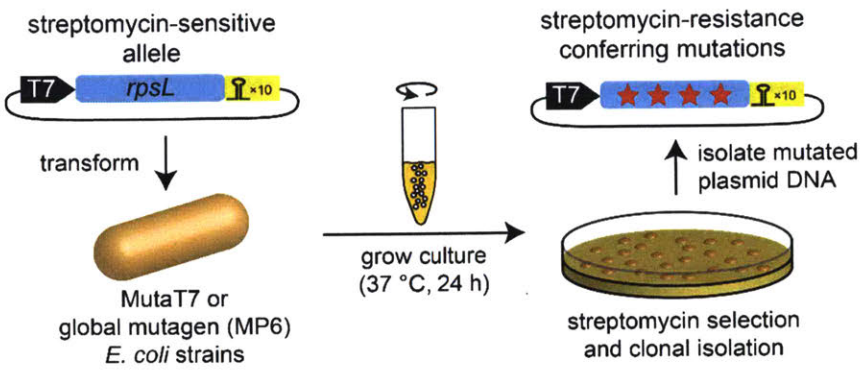

Figure 4.20 | MutaT7 introduced mutations throughout the rpsL gene

Schematic of streptomycin resistance counter-selection assay, which is designed to enrich for mutations that nullify streptomycin sensitivity. Such sensitivity is initially conferred by a streptomycin–sensitive allele of *rpsL* downstream of a T7 promoter on a reporter plasmid.

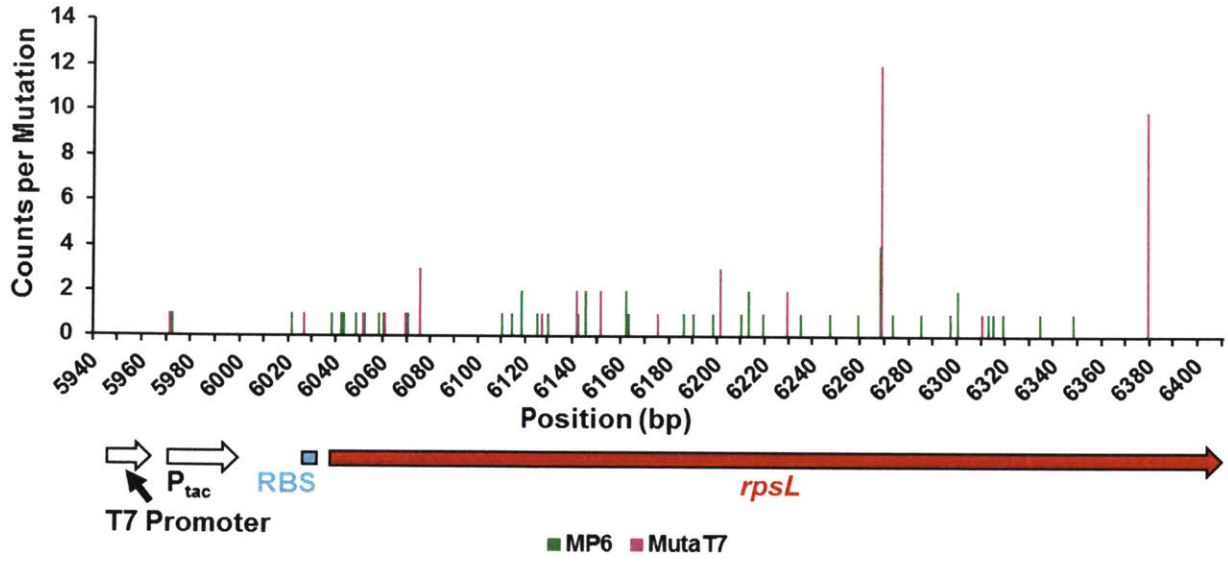

Figure 4.21 | MutaT7 introduced mutations throughout the rpsL gene

The position of various mutations throughout the T7 promoter + rpsL reporter plasmid according to the Sanger sequencing data of 48 streptomycin resistant mutants from the MP6 strain and 42 streptomycin resistant mutants from the MutaT7 strain.

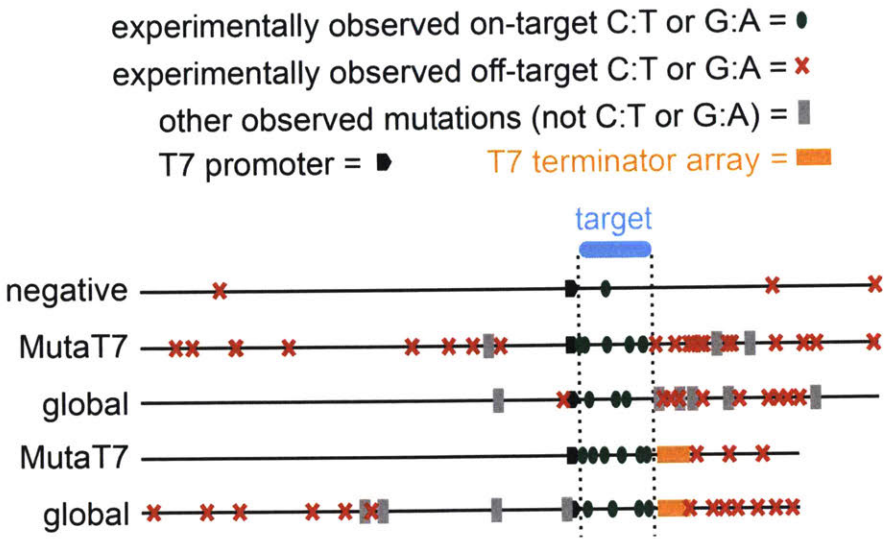

Figure 4.22 | Terminator array reduces off-target episomal mutations caused by MutaT7

Visual representation of on-target (green oval) and off-target (red x) mutations identified by sequencing episomes propagated in the presence of targeted (MutaT7) and global (MP6) mutagens.

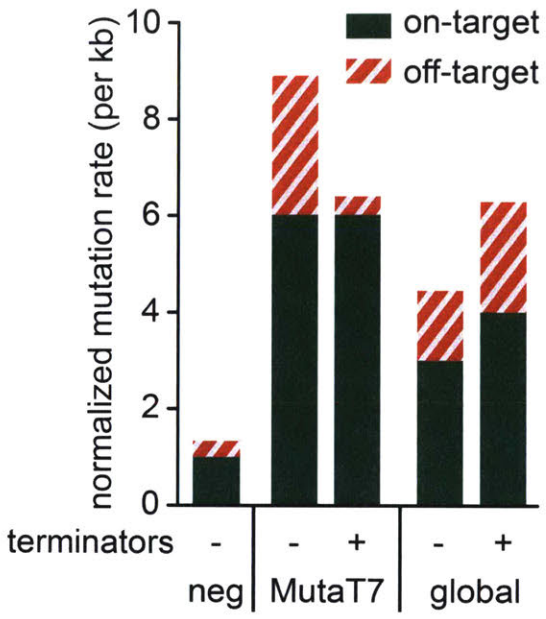

Figure 4.23 | Assessment of normalized on-target and off-target frequencies

Normalized mutation frequency (number of mutations observed divided by number of kb of DNA sequenced in associated regions) of mutations visualized in **Figure 4.22**.

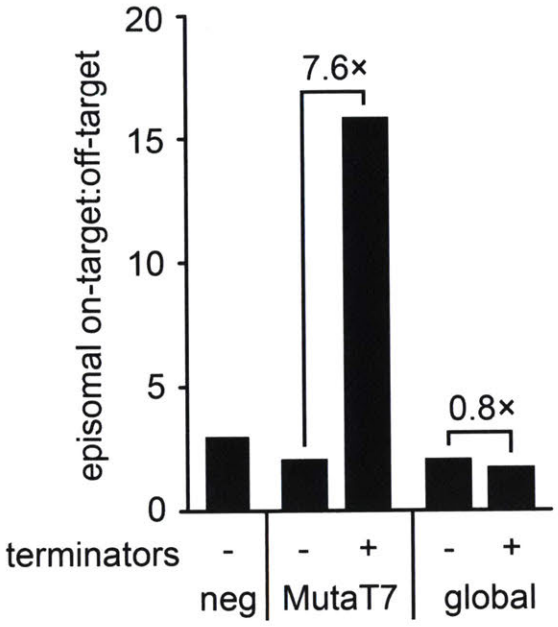

Figure 4.24 | MutaT7 preferentially mutates DNA between T7 promoter and terminator

On-target to off-target mutation ratios for data in **Figure 4.23**.

4.4.4 Expanding the mutational spectrum of mutaT7 with antisense promoters

One disadvantage of MutaT7 is its limited mutational spectrum owing to the use of a cytidine deaminase as the mutagenic component and an apparent strand bias for mutagenesis. Indeed, our sequencing results indicate that we obtain predominantly C→T transitions in the sense strand of targeted DNA using a single T7 promoter (**Figure 4.25**). We hypothesized that the mutational spectrum could be doubled by installing a second T7 promoter that would recruit MutaT7 to the 3'-end of the DNA of interest. By enabling processive activity in the opposing direction, we expected to observe a broader spectrum of mutational accumulation over time upon sampling a continuous culture of a MutaT7 strain carrying the dual promoter reporter plasmid (**Figure 4.26**). As expected, we found that the installation of an additional antisense T7 promoter leads to accumulation of both G→A and C→T mutations throughout the target gene during continuous culturing (**Figure 4.27**). Furthermore, we found that the average number and range of mutations per clone increased over time (**Figure 4.28**). This observation suggests that, in contrast to other genetic methods for global mutagenesis, where the organism often rapidly identifies a mechanism to shutdown mutagen expression, the high on-target to off-target mutation ratio of MutaT7 enables long-term maintenance of mutagen expression in cells.

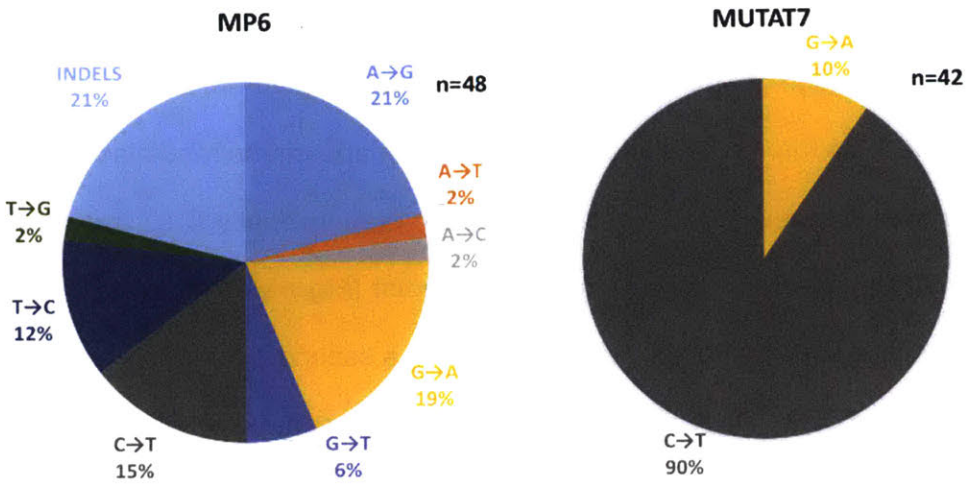

Figure 4.25 | Mutational spectra for a global mutagen compared to mutaT7

Type of mutations according to the Sanger sequencing data of 48 streptomycin resistant mutants from the MP6 strain and 42 streptomycin resistant mutants from the MutaT7 strain. Results suggest that there is significant bias for mutaT7 to introduce mutations to the sense strand (relative to the direction mutaT7 is processing).

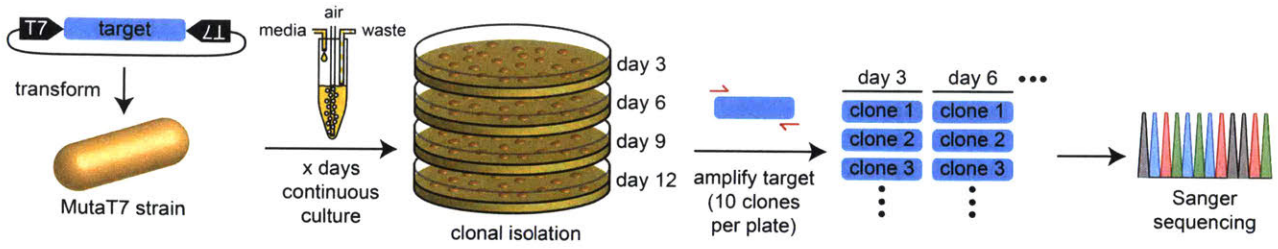

Figure 4.26 | Continuous culturing and enables sampling of mutations over time

Diagram of continuous culture conditions used to propagate a dual promoter episome in cells expressing MutaT7, along with details for downstream Sanger sequencing analysis.

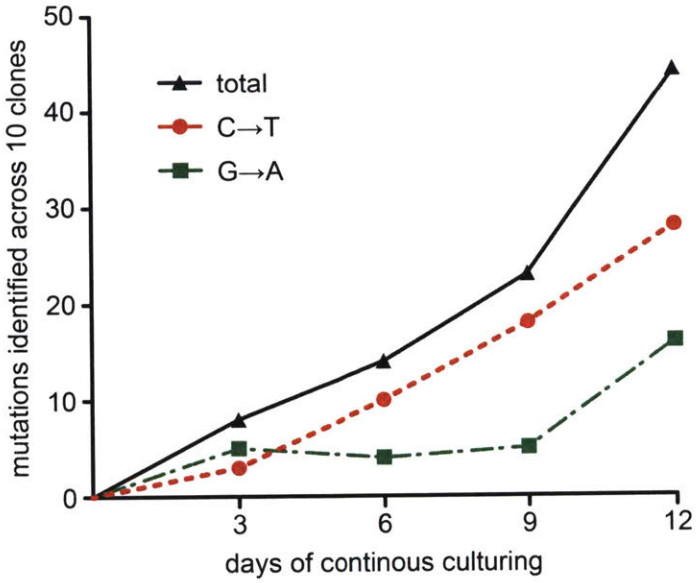

Figure 4.27 | Dual promoter architecture enables more mutation accumulation on both strands of target DNA

Graphic of mutations observed by Sanger sequencing a target gene between dual opposing T7 promoters from clones harvested at different time points (triangles for total mutations, circles for C→T transitions, and squares for G→A transitions).

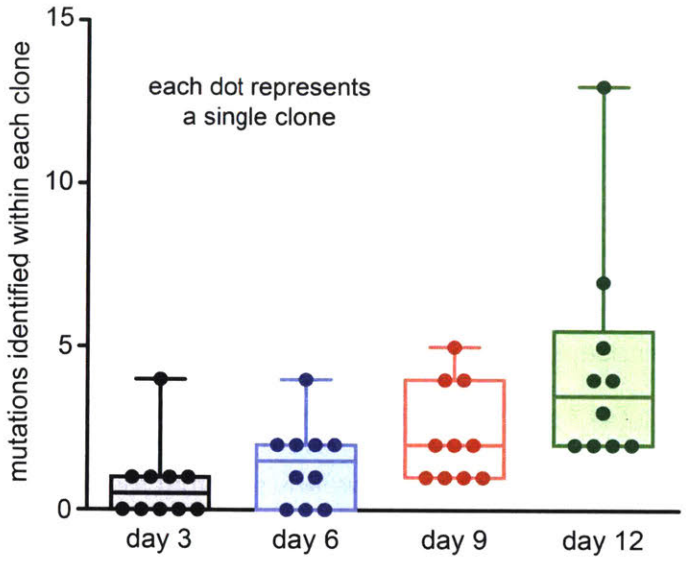

Figure 4.28 | Dual promoter architecture enables a diverse range of mutations over time

Box and whisker plot of mutations from **Figure 2.26**, where each dot represents the number of mutations found in each clone. Mean number of mutations at each time point is represented by horizontal line.

4.4.5 Concluding remarks

In summary, the processively acting MutaT7 chimera reported here is capable of selectively targeting mutations to large, yet well-defined, regions of DNA in a living system. Moreover, the availability of T7 variants with altered transcription rates[29, 30] likely provides the opportunity to fine-tune mutation rates. We anticipate that utilizing other base editing enzymes in place of cytidine deaminase, such as the adenosine deaminases[14], will significantly widen the mutational spectrum of MutaT7 and further enable the creation of rich, diverse DNA libraries *in vivo* with minimal off-target effects. Moreover, we envision that the ubiquitous applicability and high specificity of T7 RNA polymerase in a large number of diverse organisms[31-34] will enable implementation of targeted mutagenesis in a broad range of evolutionary and synthetic biology settings.

4.5 Methods

4.5.1 General

All PCR reactions for restriction cloning and recombineering targeting cassettes were performed using Q5 High Fidelity DNA Polymerase (New England Biolabs). All colony PCR reactions for sequencing were performed using OneTaq Quick-Load 2× Master Mix with Standard Buffer (New England Biolabs). Primers were obtained from Life Technologies. Gene blocks were obtained from Integrated DNA Technologies.

4.5.2 Reagents

The following reagents were obtained as indicated: Kanamycin monosulfate, fosfomycin, agar, and chloramphenicol (Alfa Aesar J61272, J6602, A10752, and B20841, respectively); tetracycline hydrochloride (CalBioChem 58346); rifampicin (TCI R0079); ampicillin (Fisher Bioreagents BP1761-25); streptomycin sulfate (MP Biomedical 100556); tetrazolium chloride, L-rhamnose, antifoam-204, and ethylmethanesulfonate (EMS) (Sigma-Aldrich T8877, W373011, A8311, and M0880, respectively); L-arabinose and cycloheximide (Chem-Impex 01654 and 00083, respectively); and lysogeny broth (LB; Difco 244620).

4.5.3 Cloning and Recombineering

All plasmids were generated by restriction cloning. Ligation reactions were performed using Quick Ligase (New England Biolabs). All DNA cloning was performed in DH10B cells (Invitrogen). The rApo1 gene was amplified from pET28b-BE1[11] and the T7 RNA polymerase gene was amplified from pTara[35]. Mutation assay reporter plasmids utilizing the single-copy BAC origin and the terminator arrays of the UUCG-T7 derivative of the T7 terminator[36] were generated by serial insertion of the annealed oligos NheI-UUCG-BamHI S and NheI-UUCG-BamHI AS (**Supplementary Table 4.3**). All *E. coli* strains used in this work were engineered using lambda red recombineering strategies.

4.5.4 Lambda Red Recombineering

The *E. coli* genome was edited using seamless lambda red recombineering with ccdB counterselection, as previously described[37]. Cells were transformed with the temperature-sensitive psc101-gbaA recombineering plasmid, plated on LB agar with 10 μg/mL tetracycline, and incubated for 24 h at 30 °C. Colonies were selected and grown in LB containing 10 μg/mL tetracycline overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 μg/mL tetracycline and grown at 30 °C for ~2 h until attaining an OD_{600} of 0.3–0.4. The ccdA antitoxin and recombineering machinery were then induced by adding arabinose and rhamnose to a final concentration of 2 mg/mL each and then growing the cultures at 37 °C for 40 min to an OD_{600} of ~0.6. The cultures were then placed on ice, washed twice with ice-cold sterile ddH_2O, resuspended in ~25 μL of ice-cold sterile ddH_2O, and electroporated with ~200 ng of the appropriate kan-ccdB targeting cassette (1.8 kV, 5.8 ms, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in super optimal broth with catabolite repression (SOC) with 2 mg/mL arabinose at 30 °C for 2 h, then plated on LB agar plates with 50 μg/mL kanamycin and 2 mg/mL arabinose and incubated for 24 h at 30 °C. Colonies that grew under these conditions had incorporated the kan-ccdB targeting cassette and were picked and grown in LB with 50 μg/mL kanamycin and 2 mg/mL arabinose at 30 °C for 18–21 h. The cultures were then diluted 25-fold in LB with 50 μg/mL kanamycin and 2 mg/mL arabinose and grown at 30 °C for ~2 h until they reached an OD_{600} of 0.3–0.4. The recombineering machinery was then induced by adding rhamnose to a final concentration of 2 mg/mL and then growing the cultures at 37 °C for 40 min to an OD_{600} of ~0.6. The cultures were then placed on ice, washed twice with ice-cold sterile ddH_2O, resuspended in ~25 μL of ice-cold sterile ddH_2O, and electroporated with ~200 ng of the final targeting cassette intended to replace the kan-ccdB cassette currently integrated in the genome (1.8 kV, 5.8 ms, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in SOC with 2 mg/mL arabinose at 30C for 2 h, and then were washed once with LB

to remove the arabinose and prevent continued production of the ccdA antitoxin. The cultures were then plated on LB agar plates at various dilutions with 100 μg/mL streptomycin and incubated for 24 h at 37 °C. Without the ccdA antitoxin, the ccdB toxin will kill cells that have not replaced the integrated kan-ccdB cassette with the final targeting cassette. The colonies that grow should have the final targeting cassette integrated, but were screened by PCR or sequencing to confirm cassette integration as some colonies may simply inactive the ccdB toxin. Once a clone with the desired change was found, the temperature-sensitive psc101-gbaA recombineering plasmid was cured by plating on LB agar with 100 μg/mL streptomycin, incubating at 42 °C for 18–21 h, streaking a colony from the plate on LB agar with 100 μg/mL streptomycin, and incubating at 42 °C for another 18–21 h. The colonies from the second plate were grown in LB with 100 μg/mL streptomycin at 37 °C to generate glycerol stocks. The colonies were also incubated in LB with 10 μg/mL tetracycline at 30 °C to ensure tetracycline sensitivity and confirm that the recombineering plasmid was successfully cured. The various strains used in this work (**Supplementary Table 4.1**) were generated using the primers in **Supplementary Table 4.3**.

The following list of modifications were made in descending order to obtain the strains used in this work:

Modification	**Genotype**	**KanccdB cassette primers used with R6K-kan-ccdB template plasmid**	**Final targeting cassette oligos or primers and template (if applicable)**	**Purpose of modification**
Insertion of rApo1 or MutaT7	Δ(*araA-leu*)*7697*::[BBa_J23114 *rApo1* or *MutaT7*]	dAraLeu7697 kanccdB F and dAraLeu7697 kanccdB R	dAraLeu7697-rApol and dAraLeu7697-T7 used to amplify from BBa_J23114 lacO rApo1 or BBa_J23114_lacO MutaT7	rApo1 and MutaT7 were inserted into the DH10B genome between basepairs 62,378 and 62,379 at the seam of the large Δ(*araA-leu*)*7697* deletion[38].
ung Deletion	Δ*ung*	5'-Ung kanccdB and 3'-Ung kanccdB	oligos delUng S and delUng AS (annealed oligos)	To minimize dU→dC repair, uracil DNA glycosylase (*ung*) was deleted[39].
Increasing *lacI* expression	[P_{lacI}<>P_{tac}]	5'-pLacI::kanccdB and 3'-pLacI::kanccdB	pLacI::pTac S and pLacI::pTac AS (annealed oligos)	In an attempt to get very low basal expression from lacI repressed promoters, the endogenous P_{lacI} promoter was replaced with the strong P_{tac} promoter[24]
Replacement of promoter BBa_J23114 with $P_{A1lacO-Tenth}$	Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *rApo1* or *MutaT7*]	5'-prApol::kanccdB and 3'-prApol::kanccdB	PA1lacO-1 F and PA1lacO-1 R used to amplify from pA1lacO-tenth gene block (**Supplementary Table 4.3**)	The BBa_J23114 promoter from the Anderson Collection[23] that controlled the expression of rApo1 or MutaT7 from the DH10B genome was replaced with the tightly-repressed promoter $P_{A1lacO-Tenth}$, which is a weaker version of the P_{A1lacO} promoter[22].
Deactivation of rApol	Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ d*rApo1* or *drApo1-T7*]	5'-drApol::kanccdB and 3'-drApol::kanccdB	drApol S and drApol AS (annealed oligos)	The E63Q mutant of rApo1 cytidine deaminase has been shown to be catalytically inactive[21]. E63Q mutant negative control strains were made.

4.5.5 Deleting the motAB and csgABCDEFG operons through DIRex lambda red recombineering to decrease biofilm formation in bioreactor experiments

Deletions of the *motAB* operon[40] and the *csgABCDEFG*[41] have been shown to produce strains of *E. coli* that are deficient in biofilm formation. To minimize inlet line contamination and clogs in bioreactor experiments owing to biofilms, the *motAB* and *csgABCDEFG* operons were deleted using one-step DIRex lambda red recombineering[42]. The *motAB* targeting half-cassettes were amplified from R6K-AmilCP-kan-ccdB using the primers delmotDF and AmilCP-KanR and from R6K-kan-ccdB-AmilCP using the primers delmotDR and KanF-AmilCP (**Supplementary Table 4.3**). The *motAB* half cassettes were co-electroporated to replace *motAB* with a kan-ccdB cassette flanked by large AmilCP inverted repeats nested between short 30 bp direct repeats. The repeat architecture leads to a high rate of spontaneous excision that was selected for using ccdB counterselection to obtain a markerless deletion of *motAB*. This procedure was then repeated to delete the *csgABCDEFG* operon. The *csgABCDEFG* targeting half-cassettes were amplified from R6K-AmilCP-kan-ccdB using the primers delcsgDF and AmilCP-KanR and from R6K-kan-ccdB-AmilCP using the primers delcsgDR and KanF-AmilCP (**Supplementary Table 4.3**).

4.5.6 Mutation Assay

To assess mutagenesis rates, the control (Δ*ung*, rApo1, drApo1, and drApo1–T7; **Supplementary Table 4.1**) and mutagenic strains (MutaT7 and MP6; **Supplementary Table 4.1**) (Strep^R) carrying reporter plasmids (Amp^R) were streaked on LB agar with 100 μg/mL streptomycin and 100 μg/mL ampicillin and grown at 37 °C for 24 h in order to obtain clones. Single colonies were picked in triplicate for each sample and used to inoculate 5 mL LB with 100 μg/mL streptomycin,100 μg/mL ampicillin, and 25 mM arabinose (with 10 μg/mL chloramphenicol for the MP6 strain, **Supplementary Table 4.1**), then shaken at 250 r.p.m. and 37 °C for 24 h to accumulate mutations during growth. 1 mL aliquots of each culture were

pelleted at 6000 × *g* for 3 min and resuspended in 1 mL LB to remove arabinose. Each resuspension was plated on LB agar plates with 50 μg/mL tetrazolium chloride (a metabolic contrast dye for visualizing colonies) and the antibiotics indicated below to analyze mutations rates and viability:

- 50 μL of a 100,000-fold dilution of each resuspension was plated on LB agar with 100 μg/mL streptomycin, 100 μg/mL ampicillin, and 50 μg/mL tetrazolium chloride. For samples from the MP6 strain, owing to lower growth of that strain, 50 μL of a 10,000-fold dilution of each resuspension was plated to obtain a more accurate count. The colony counts from these plates were used to calculate the cell viability (i.e., the number of live, ampicillin resistant cells) in CFU/mL for each sample (**Supplementary Tables 4.2a–c** and **Figure 4.15**).
- 50 μL of each resuspension was plated on LB agar plates with 200 μg/mL kanamycin and 50 μg/mL tetrazolium chloride. The colony counts from these plates were used to calculate the number of kanamycin resistant mutants in CFU/mL for each sample (**Supplementary Tables 4.2a–c** and **Figure 4.17**). The number of kanamycin resistant mutants in CFU/mL was divided by the number of live ampicillin resistant cells in CFU/mL for each sample to obtain the kanamycin resistant mutation frequency (**Supplementary Tables 4.2a–c** and **Figure 4.10**).
- 50 μL of each resuspension was plated on LB agar plates with 20 μg/mL tetracycline and 50 μg/mL tetrazolium chloride. The colony counts from these plates were used to calculate the number of tetracycline resistant mutants in CFU/mL for each sample (**Supplementary Tables 4.2a–c**). The number of tetracycline resistant mutants in CFU/mL was divided by the number of live ampicillin resistant cells in CFU/mL for each sample to obtain the tetracycline resistant mutation frequency (**Supplementary Tables 4.2a–c** and **Figure 4.10**).
- 50 μL of each resuspension was plated on LB agar plates with 100 μg/mL rifampicin and 50 μg/mL tetrazolium chloride. The colony counts from these plates were used to calculate the

number of rifampicin resistant mutants in CFU/mL for each sample (**Supplementary Tables 4.2a–c**). The number of rifampicin resistant mutants in CFU/mL was divided by the number of live ampicillin resistant cells in CFU/mL for each sample to obtain the rifampicin resistant mutation frequency (**Supplementary Tables 4.2a–c** and **Figure 4.12**).

- 50 μL of each resuspension was plated on LB agar plates with 100 μg/mL fosfomycin and 50 μg/mL tetrazolium chloride. The colony counts from these plates were used to calculate the number of fosfomycin resistant mutants in CFU/mL for each sample (**Supplementary Tables 4.2a–c**). The number of fosfomycin resistant mutants in CFU/mL was divided by the number of live ampicillin resistant cells in CFU/mL for each sample to obtain the rifampicin resistant mutation frequency (**Supplementary Tables 4.2a–c** and **Figure 4.14**).

Plates were incubated at 37 °C for 48 h, then imaged by inverting the plates onto transparencies and scanning on a document scanner at a resolution of 400 dots per inch. The colonies were then counted using the software OpenCFU (3.9.0)[25], with the minimum colony radius set to 3, the maximum colony radius set to 50, and the regular threshold set to 4.

4.5.7 Chemical Mutagenesis with Chemical Mutagen EMS

Mutagenesis with EMS was performed as previously described[5]. An overnight culture of each sample was subcultured and grown until it reached a density of 2–3 × 10^8 cells per mL (log phase). 5 mL aliquots of cells were chilled on ice, washed twice with sodium phosphate buffer (pH = 7), and resuspended in 1 mL of 1× PBS in a 1.5 mL Eppendorf tube. EMS was added while cold by pipetting 14 μL of EMS into 1 ml of resuspended cells. Eppendorfs were sealed and mixed at 1000 r.p.m. for 60 min at 37 °C. The cells were then washed twice with LB and resuspended in 1 mL of LB. Immediately after washing, a viability measurement was performed by plating 50 μL of a 10,000-fold dilution of each culture on LB agar with 100 μg/mL streptomycin, 100 μg/mL ampicillin, and 50 μg/mL tetrazolium chloride. After 48 h of incubation, plates were imaged on a document scanner as described above. The number of live ampicillin

colonies were counted after EMS treatment in CFU/mL to measure the viability after mutagen treatment (**Supplementary Tables 4.2a–c** and **Figure 4.15**). For mutation rate assessment, 500 μL of the post-EMS-treated resuspension was inoculated into 5 ml of LB with 100 μg/mL streptomycin and 100 μg/mL ampicillin. The cultures were grown at 37 °C for 20 h, then 50 μL of each culture was plated on LB agar with 50 μg/mL tetrazolium chloride and 100 μg/mL rifampicin. 50 μL of a 100,000-fold dilution of each culture was also plated on LB agar with 100 μg/mL streptomycin, 100 μg/mL ampicillin, and 50 μg/mL tetrazolium chloride. After 48 h of incubation, plates were imaged on a document scanner as described above. The number of rifampicin resistant mutants in CFU/mL was divided by the number of live ampicillin resistant cells in CFU/mL for each sample to obtain the rifampicin resistant mutation frequency (**Supplementary Tables 4.2a–c** and **Figure 4.12**).

4.5.8 Mutation Assay and Sequencing with the T7 Promoter + rpsL Reporter Plasmid

To assess the locations and types of mutations observed, the drApo1-T7 negative control strain and MutaT7 and MP6 mutagenic strains (**Supplementary Table 4.1**) ($Strep^R$) carrying the T7 promoter +rpsL reporter plasmid (Amp^R) were streaked on LB agar with 100 μg/mL ampicillin and grown at 37 °C for 24 h in order to obtain clones. Single colonies were picked in triplicate for each sample and used to inoculate 5 mL LB with 100 μg/mL ampicillin and 25 mM arabinose (with 10 μg/mL chloramphenicol for the MP6 strain, **Supplementary Table 4.1)**, then shaken at 250 r.p.m. and 37 °C for 24 h to accumulate mutations during growth. 1 mL aliquots of each culture were pelleted at 6000 × *g* for 3 min and resuspended in 1 mL LB to remove arabinose. 50 μL of a 100-fold dilution of each resuspension was plated on LB Lennox agar plates (pH 8.0) with 500 μg/mL streptomycin, 100 μg/mL ampicillin, and 50 μg/mL tetrazolium chloride. 48 colonies from each plate were picked for colony PCR using the primers 2062 and 1197 (**Supplementary Table 4.3**). The amplicons were Sanger-sequenced using the primer 1197 (**Supplementary Table 4.3**).

4.5.9 Continuous Culturing and Sequencing of the Dual T7 Promoter Reporter Plasmid

The dual T7 promoter reporter plasmid was continuously cultured in the MutaT7-*csg*$^{+}$ *mot*$^{+}$ strain (**Supplementary Table 4.1**) in a 70 mL culture in a round-bottomed flask that was slowly stirred in a 37 °C mineral oil bath. The culture was aerated through a needle that was connected to a standard aquarium pump and LB with 100 μg/mL streptomycin, 100 μg/mL ampicillin, and 0.5% isopropanol (as an antifoaming agent) was fed into the culture via a needle connected to a peristaltic pump at a rate of ~0.5 volumes/h. Fractions were collected every 3 d for 12 d. Each fraction was plated for single colonies on LB agar with 100 μg/mL ampicillin and 10 clones from each fraction were Sanger-sequenced by colony PCR with the primers 1493 and 1494 (**Supplementary Table 4.3**).

4.5.10 Continuous Culturing and Sequencing of the T7 Promoter + Filler DNA and T7 Promoter + Terminators Reporter Plasmids Reporter Plasmids

The T7 promoter + filler DNA and T7 promoter + terminators reporter plasmids were continuously cultured in the Δ*ung* (negative control), MutaT7, and MP6 strains (**Supplementary Table 4.1**) in 20 mL cultures using a previously described multiplex bioreactor setup[43]. The reactor was stored in a 37 °C warm room and was aerated and stirred with aquarium pumps. LB with 100 μg/mL streptomycin, 100 μg/mL ampicillin, 100 μg/mL cycloheximide, 0.01% (v/v) antifoam-204, and 150 μg/mL arabinose (+10 μg/mL chloramphenicol in the case of the MP6 strain (**Supplementary Table 4.1**)) was pumped into each reaction vessel at a rate of 0.87 volumes/h. Fractions were collected every 3 d. Each fraction was plated on LB agar with 100 μg/mL streptomycin and 100 μg/mL ampicillin and 12 single colonies from each plate were grown in 5 mL LB with 100 μg/mL ampicillin. DNA was isolated from each overnight culture using the Qiaprep 96 Turbo Miniprep Kit and quantified using the PicoGreen assay.

4.5.11 Library Construction and Next Generation Sequencing

Libraries were prepared using a miniaturized version of Nextera XT. Briefly, 0.5 ng of input DNA was subjected to a 1/12 scale reaction of Illumina Nextera XT performed on a TTP Labtech Mosquito HV using combinatorial dual indexing (V_{final} = 4 μl). Completed libraries were size selected using SPRI beads at 0.7× volume and pooled before sequencing on an Illumina MiSeq using 150 nt paired end reads (v2 chemistry). Sequencing reads were aligned against respective plasmid sequences using bwa mem 0.7.10-r789 [RRID:SCR_010910]. Allele pileups were generated using samtools v.0.1.19 mpileup [RRID:SCR_002105] with flags -d 10000000 --excl-flags 2052, and allele counts/frequencies were extracted[44, 45]. Only positions with greater than 10-fold coverage in all replicates of each sample were included in the analysis. Fixed variant alleles (present at greater than 85% frequency) for each sample are reported. Sanger sequencing was also performed on a PCR amplicon from 96 clones of Δ*ung* (negative control) and MutaT7 strains (**Supplementary Table 4.1**) after 15 d of continuous culture carrying the T7 promoter + terminators reporter plasmid. The primers 2165 and 1197 (**Supplementary Table 4.3**) were used to amplify and Sanger sequence the Kan^R gene.

4.6 Supplementary Tables

Supplementary Table 4.1 | Strain table

The genotypes of strains used in this work are shown. The "x<>y" notation indicates a replacement of "x" with "y" through lambda red recombineering.

Strain	Genotype
DH10B[38]	F^- *araD139* Δ(*araA-leu*)*7697* Δ*lacX74 galE15 galK16 galU hsdR2 relA1 rpsL150*(Str^R) *spoT1* φ80*lacZ*ΔM15 *endA1 nupG recA1 e14*$^-$ *mcrA* Δ(*mrr hsdRMS-mcrBC*)
Δ*ung*	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG*
rApo1	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO\text{-}Tenth}$ *rApo1*]
drApo1	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO\text{-}Tenth}$ *drApo1*]
MutaT7	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO\text{-}Tenth}$ *MutaT7*]
drApo1–T7	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO\text{-}Tenth}$ *drApo1-T7*]
MP6	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* MP6(Cm^R)
MutaT7-*csg*$^+$ *mot*$^+$	DH10B Δ*ung* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO\text{-}Tenth}$ *MutaT7*]

Supplementary Table 4.2a | Drug resistance assay data

Strain	Reporter Plasmid	Resistance	CFU/mL	Standard Deviation	Resistance Frequency	Standard Deviation
drApo1	No T7 promoter + filler DNA	Kan	6.0E+1	4.0E+1	3.8E–8	2.6E–8
		Tet	1.8E+2	9.2E+1	1.1E–7	6.3E–8
		Fos	3.3E+3	5.3E+2	2.0E–6	4.2E–7
		Rif	7.6E+2	1.0E+3	4.5E–7	5.9E–7
		Amp	1.6E+9	1.0E+8		
	T7 promoter + filler DNA	Kan	6.5E+2	3.9E+2	5.2E–7	3.4E–7
		Tet	6.3E+2	3.9E+2	4.6E–7	2.3E–7
		Fos	3.5E+3	4.2E+1	2.7E–6	5.8E–7
		Rif	4.0E+2	1.1E+2	3.0E–7	3.0E–8
		Amp	1.3E+9	2.5E+8		
	T7 promoter + terminator array	Kan	1.2E+3	1.2E+3	6.6E–7	6.1E–7
		Tet	3.7E+2	2.3E+1	2.1E–7	3.6E–8
		Fos	3.4E+3	1.3E+3	1.9E–6	7.4E–7
		Rif	5.9E+2	4.2E+2	3.4E–7	2.7E–7
		Amp	1.8E+9	2.0E+8		
Δung	No T7 promoter + filler DNA	Kan	4.1E+2	1.2E+2	2.6E–7	9.8E–8
		Tet	8.1E+2	8.6E+2	4.9E–7	4.8E–7
		Fos	3.3E+3	5.2E+2	2.1E–6	3.0E–7
		Rif	2.3E+2	8.1E+1	1.4E–7	3.8E–8
		Amp	1.6E+9	1.4E+8		
	T7 promoter + filler DNA	Kan	6.9E+2	2.5E+2	3.8E–7	1.5E–7
		Tet	6.3E+2	2.3E+1	3.4E–7	2.1E–8
		Fos	4.5E+3	6.6E+2	2.5E–6	4.9E–7
		Rif	3.0E+2	2.0E+1	1.7E–7	2.0E–8
		Amp	1.8E+9	1.2E+8		
	T7 promoter + terminator array	Kan	5.9E+2	4.8E+2	6.7E–7	8.2E–7
		Tet	9.2E+2	7.8E+2	1.0E–6	1.3E–6
		Fos	3.7E+3	1.0E+3	3.6E–6	2.8E–6
		Rif	2.0E+2	1.0E+2	1.9E–7	1.6E–7
		Amp	1.3E+9	5.6E+8		

Supplementary Table 4.2b | Drug resistance assay data

Strain	Reporter Plasmid	Resistance	CFU/mL	Standard Deviation	Resistance Frequency	Standard Deviation
drApo1–T7	No T7 promoter + filler DNA	Kan	1.1E+2	4.2E+1	1.2E–7	3.3E–8
		Tet	2.2E+2	7.2E+1	2.6E–7	6.1E–8
		Fos	2.6E+3	2.3E+2	3.3E–6	1.0E–6
		Rif	1.7E+2	1.2E+2	2.2E–7	1.8E–7
		Amp	8.7E+8	3.2E+8		
	T7 promoter + filler DNA	Kan	2.3E+2	1.2E+2	1.2E–7	3.9E–8
		Tet	5.9E+2	3.4E+2	3.9E–7	3.2E–7
		Fos	3.0E+3	2.1E+2	1.8E–6	6.5E–7
		Rif	2.6E+2	8.7E+1	1.5E–7	3.8E–8
		Amp	1.8E+9	6.5E+8		
	T7 promoter + terminator array	Kan	1.5E+2	5.0E+1	1.3E–7	8.2E–8
		Tet	3.7E+2	1.2E+1	3.1E–7	9.8E–8
		Fos	3.3E+3	7.9E+2	2.6E–6	5.7E–7
		Rif	1.6E+2	2.0E+1	1.3E–7	2.3E–8
		Amp	1.3E+9	3.2E+8		
MutaT7	No T7 promoter + filler DNA	Kan	1.7E+2	7.6E+1	1.2E–7	6.1E–8
		Tet	2.5E+2	2.4E+2	1.6E–7	1.4E–7
		Fos	2.8E+3	8.7E+2	1.9E–6	3.8E–7
		Rif	5.5E+2	1.0E+2	3.7E–7	5.2E–8
		Amp	1.5E+9	1.8E+8		
	T7 promoter + filler DNA	Kan	1.1E+4	3.1E+3	7.2E–6	2.9E–6
		Tet	3.1E+4	2.3E+3	2.0E–5	4.2E–6
		Fos	2.5E+3	1.1E+3	1.6E–6	6.3E–7
		Rif	4.2E+2	1.2E+2	2.6E–7	1.5E–8
		Amp	1.6E+9	3.6E+8		
	T7 promoter + terminator array	Kan	9.8E+3	1.9E+3	6.1E–6	5.5E–7
		Tet	5.0E+2	2.4E+2	3.0E–7	1.2E–7
		Fos	3.7E+3	6.3E+2	2.3E–6	2.4E–7
		Rif	9.1E+2	5.1E+2	5.6E–7	2.7E–7
		Amp	1.6E+9	2.1E+8		

Supplementary Table 4.2c | Drug resistance assay data

Strain	Reporter Plasmid	Resistance	CFU/mL	Standard Deviation	Resistance Frequency	Standard Deviation
MP6	No T7 promoter + filler DNA	Kan	2.9E+3	4.7E+2	4.8E–5	1.6E–6
		Tet	3.5E+3	5.6E+2	5.9E–5	8.5E–6
		Fos	1.5E+4	1.9E+3	2.5E–4	3.9E–5
		Rif	9.5E+3	2.9E+3	1.5E–4	3.1E–5
		Amp	6.0E+7	8.2E+6		
	T7 promoter + filler DNA	Kan	3.3E+3	4.6E+2	8.0E–5	2.8E–5
		Tet	4.8E+3	7.8E+2	1.1E–4	1.2E–5
		Fos	1.2E+4	1.7E+3	2.9E–4	3.5E–5
		Rif	9.3E+3	1.7E+3	2.2E–4	4.9E–5
		Amp	4.4E+7	1.2E+7		
	T7 promoter + terminator array	Kan	9.9E+3	1.2E+3	3.4E–5	1.0E–5
		Tet	1.5E+4	3.6E+3	5.3E–5	2.0E–5
		Fos	2.2E+4	1.3E+3	7.7E–5	1.8E–5
		Rif	2.2E+4	3.4E+3	7.5E–5	2.5E–5
		Amp	3.0E+8	6.6E+7		
EMS	No T7 promoter + filler DNA	Rif	2.7E+4	1.6E+4	3.1E–5	1.4E–5
		Amp	9.3E+8	4.3E+8		
		Amp*	5.3E+5	7.6E+5		
	T7 promoter + filler DNA	Rif	3.6E+4	3.9E+3	3.7E–5	6.5E–6
		Amp	9.6E+8	9.1E+7		
		Amp*	1.3E+6	1.0E+6		
	T7 promoter + terminator array	Rif	2.4E+4	5.5E+3	4.9E–5	1.7E–5
		Amp	5.4E+8	2.5E+8		
		Amp*	1.4E+6	6.0E+5		

*The number of ampicillin resistant cells immediately after EMS treatment.

Supplementary Table 4.3 | Primer and oligo table

Primer Name	SEQUENCE
5'–Ung kanccdB	GCAGTTAAGCTAGGCGGATTGAAGATTCGCAGGAGAGCGAGATGGCTAACCCCTCATCAGTGCCAACAT AGTAAG
3'–Ung kanccdB	AGCCGGGTGGCAACTCTGCCATCCGGCATTTCCCCGCAAATTTACTCACTCCGCTCATTAGGCGGGC
delUng S	GCAGTTAAGCTAGGCGGATTGAAGATTCGCAGGAGAGCGAGATGGCTAACAGTGAGTAAATTTGCGGG GAAATGCCGGATGGCAGAGTTGCCACCCGGCT
delUng AS	AGCCGGGTGGCAACTCTGCCATCCGGCATTTCCCCGCAAATTTACTCACTGTTAGCCATCTCGCTCTCC TGCGAATCTTCAATCCGCCTAGCTTAACTGC
5'–pLacI::kanccdB	CGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTCCCTCATCAGTGCCAACATA GTAAG
3'–pLacI::kanccdB	TGGTGGCCGGAAGGCGAAGCGGCATGCATTTACGTTGACACCATCGAATGCCGCTCATTAGGCGGGC
pLacI::pTac AS	ATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTCATTATACGAGCCGATGATTAATTGTCAACATTCG ATGGTGTCAACGTAAATGCATGCCGCTTC
pLacI::pTac S	GAAGCGGCATGCATTTACGTTGACACCATCGAATGTTGACAATTAATCATCGGCTCGTATAATGAGCGCC CGGAAGAGAGTCAATTCAGGGTGGTGAAT
delcsgDF	TTTCGCTTAAACAGTAAAATGCCGGATGATAATTCCGGCTTTTTTATCTGTTTGTTGAAATATCGGAATTA AAAAAGAATTCAAAAAAAAGCCCGC
delcsgDR	GCAGCAGACCATTCTCTCCAGATTCATCTTATGCTCGATATTTCAACAAACAGATAAAAAAGCCGGAATTA AAAAAGAATTCAAAAAAAAGCCCGC
delmotDF	CTTCATCAAAAAATGTCTGATAAAAATCGCTTATATCCATGCTCACGCTGGACATCATCCTTCCAGAATTA AAAAAGAATTCAAAAAAAAGCCCGC
delmotDR	CGCCTGACGACTGAACATCCTGTCATGGTCAACAGTGGAAGGATGATGTCCAGCGTGAGCATGGAGAAT TAAAAAAGAATTCAAAAAAAAGCCCGC
KanF–AmilCP	GCTCGACGTTGTCACTGAAGC
AmilCP–KanR	CGCCGTCGGGCATGC
5'–drApoI::kanccdB	GTCGTCACTCTATCTGGCGTCACACCTCTCAGAACACCAACAAACACGTTCCGCTCATTAGGCGGGC
3'–drApoI::kanccdB	TTCGGGCAGAAGTAACGTTCGGTGGTGAATTTTTCGATGAAGTTAACTTCCCTCATCAGTGCCAACATAG TAAG
drApoI S	GTCGTCACTCTATCTGGCGTCACACCTCTCAGAACACCAACAAACACGTTCAAGTTAACTTCATCGAAAA ATTCACCACCGAACGTTACTTCTGCCCGAA
drApoI AS	TTCGGGCAGAAGTAACGTTCGGTGGTGAATTTTTCGATGAAGTTAACTTGAACGTGTTTGTTGGTGTTCT GAGAGGTGTGACGCCAGATAGAGTGACGAC
dAraLeu7697 kanccdB F	CAGCAGCAGAACGCCCGGCACGTGCTCTGCCAGTTGTTCAATGGCCTGATCCCTCATCAGTGCCAACAT AGTAAG
dAraLeu7697 kanccdB R	TGAGCAGGCAATCAGCAGTTGATAACCCCGTTGCCGCGCCTGGCGTTCAACCGCTCATTAGGCGGGC
dAraLeu7697–rApoI	CAGCAGCAGAACGCCCGGCACGTGCTCTGCCAGTTGTTCAATGGCCTGATTCGAGTTTATGGCTAGCTC AGTCC
dAraLeu7697–T7	TGAGCAGGCAATCAGCAGTTGATAACCCCGTTGCCGCGCCTGGCGTTCAACTGAAAATCTTCTCTCATC CGCC
5'–prApoI::kanccdB	GCAGAACGCCCGGCACGTGCTCTGCCAGTTGTTCAATGGCCTGATTCGAGCCGCTCATTAGGCGGGC
3'–prApoI::kanccdB	CGACGACGCAGGGTCGGGTCAACCGCAACCGGACCGGTTTCAGAAGACATCCCTCATCAGTGCCAACA TAGTAAG
PA1lacO–1 F	GCAGAACGCCCGGCAC
PA1lacO–1 R	CGACGCAGGGTCGGG
pA1lacO1–HA Tenth gene block	GCAGAACGCCCGGCACGTGCTCTGCCAGTTGTTCAATGGCCTGATTCGAGAAAGAGTGTTTATTTGTGA GCGGATAACAATTACAATTAGATTCAATTGTGAGCGGATAACAATTTCACACAGGCTAGCGAATTCGAGC TCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCTACCCATACGACG TACCAGATTACGCTATGTCTTCTGAAACCGGTCCGGTTGCGGTTGACCCGACCCTGCGTCGTCG
NheI–UUCG–BamHI S	CTAGCCAGCTTGGGTCTCCCTAGGTCGAGCTCCGTCGACCTAGCATAACCCCGCGGGGCCTCTTCGGG GGTCTCGCGGGGTTTTTTGCTGAAAGG
NheI–UUCG–BamHI AS	GATCCCTTTCAGCAAAAAACCCCGCGAGACCCCCGAAGAGGCCCCGCGGGGTTATGCTAGGTCGACGG AGCTCGACCTAGGGAGACCCAAGCTGG
1493	AAAAAAAAGCTTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGA ACTAAACCACGGGATCGGCCATTGAAC
1494	AAAAAATTAATTAAGCTCTAGAGAATTGATCCCCTCAG
2165	TTGACAATTAATCATCGGCTCGAAGCTTG
1197	AAAAAAGGATCCTTCGCCATTCAGGCTGCG
2062	AAAGTGCCACCTGGCGG

4.7 References

1. Wong, T., Zhurina, D. & Schwaneberg, U. The Diversity Challenge in Directed Protein Evolution. *Comb. Chem. High Throughput Screen.* **9**, 271-288 (2006).
2. Badran, A.H. & Liu, D.R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
3. Zhao, H. & Arnold, F.H. Combinatorial protein design: strategies for screening protein libraries. *Curr. Opin. Struct. Biol.* **7**, 480-485 (1997).
4. Tizei, P.A., Csibra, E., Torres, L. & Pinheiro, V.B. Selection platforms for directed evolution in synthetic biology. *Biochem. Soc. Trans.* **44**, 1165-1175 (2016).
5. Cupples, C.G. & Miller, J.H. A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. *Proc. Natl. Acad. Sci.* **86**, 5345-5349 (1989).
6. Tessman, I., Ishiwa, H. & Kumar, S. Mutagenic Effects of Hydroxylamine in vivo. *Science* **148**, 507-508 (1965).
7. Greener, A., Callahan, M. & Jerpseth, B. An efficient random mutagenesis technique using an E. coli mutator strain. *Mol. Biotechnol.* **7**, 189-195 (1997).
8. Gerdes, S.Y. et al. Experimental Determination and System Level Analysis of Essential Genes in Escherichia coli MG1655. *J. Bacteriol.* **185**, 5673-5684 (2003).
9. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096-1101 (2015).
10. Hong, J.S. & Ames, B.N. Localized mutagenesis of any specific small region of the bacterial chromosome. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 3158-3162 (1971).
11. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
12. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353** (2016).
13. Komor, A.C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
14. Gaudelli, N.M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).
15. Hess, G.T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036-1042 (2016).
16. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029-1035 (2016).
17. Rong, M., He, B., McAllister, W.T. & Durbin, R.K. Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 515-519 (1998).
18. Thiel, V., Herold, J., Schelle, B. & Siddell, S.G. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **82**, 1273-1281 (2001).

19. Ramiro, A.R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M.C. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat. Immunol.* **4**, 452-456 (2003).
20. Lykke-Andersen, J. & Christiansen, J. The C-terminal carboxy group of T7 RNA polymerase ensures efficient magnesium ion-dependent catalysis. *Nucleic Acids Res.* **26**, 5630-5635 (1998).
21. Navaratnam, N. et al. Evolutionary origins of apoB mRNA editing: Catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site. *Cell* **81**, 187-195 (1995).
22. Camsund, D., Heidorn, T. & Lindblad, P. Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium. *J. Biol. Eng.* **8** (2014).
23. Anderson, J.C. Anderson Promoter Collection. Available: http://parts.igem.org/Promoters/Catalog/Anderson.
24. Glascock, C.B. & Weickert, M.J. Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in Escherichia coli. *Gene* **223**, 221-231 (1998).
25. Geissmann, Q. OpenCFU, a new free and open-source software to count cell colonies and other circular objects. *PLoS One* **8**, e54072 (2013).
26. Hecht, A. et al. Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.* **45**, 3615-3626 (2017).
27. Garibyan, L. Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. *DNA Repair* **2**, 593-608 (2003).
28. Nilsson, A.I., Berg, O.G., Aspevall, O., Kahlmeter, G. & Andersson, D.I. Biological Costs and Mechanisms of Fosfomycin Resistance in Escherichia coli. *Antimicrob. Agents Chemother.* **47**, 2850-2858 (2003).
29. Bonner, G., Lafer, E.M. & Sousa, R. Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.* **269**, 25120-25128 (1994).
30. Guillerez, J., Lopez, P.J., Proux, F., Launay, H. & Dreyfus, M. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5958-5963 (2005).
31. McBride, K.E., Schaaf, D.J., Daley, M. & Stalker, D.M. Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7301-7305 (1994).
32. Lieber, A., Sandig, V. & Strauss, M. A mutant T7 phage promoter is specifically transcribed by T7-RNA polymerase in mammalian cells. *Eur. J. Biochem.* **217**, 387-394 (1993).
33. Weinstock, M.T., Hesek, E.D., Wilson, C.M. & Gibson, D.G. Vibrio natriegens as a fast-growing host for molecular biology. *Nat. Methods* **13**, 849-851 (2016).
34. Dower, K. & Rosbash, M. T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA* **8**, 686-697 (2002).
35. Wycuff, D.R. & Matthews, K.S. Generation of an AraC-araBAD promoter-regulated T7 expression system. *Anal. Biochem.* **277**, 67-73 (2000).
36. Mairhofer, J., Wittwer, A., Cserjan-Puschmann, M. & Striedner, G. Preventing T7 RNA polymerase read-through transcription-A synthetic termination signal capable of improving bioprocess stability. *ACS Synth. Biol.* **4**, 265-273 (2015).

37. Wang, H. et al. Improved seamless mutagenesis by recombineering using ccdB for counterselection. *Nucleic Acids Res.* **42**, e37 (2014).
38. Durfee, T. et al. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190**, 2597-2606 (2008).
39. Duncan, B.K. Isolation of insertion, deletion, and nonsense mutations of the uracil-DNA glycosylase (ung) gene of Escherichia coli K-12. *J. Bacteriol.* **164**, 689-695 (1985).
40. Pratt, L.A. & Kolter, R. Genetic analysis of Escherichia colibiofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol. Microbiol.* **30**, 285-293 (1998).
41. Prigent-Combaret, C. et al. Developmental pathway for biofilm formation in curli-producing Escherichia coli strains: role of flagella, curli and colanic acid. *Environ. Microbiol.* **2**, 450-464 (2000).
42. Nasvall, J. Direct and Inverted Repeat stimulated excision (DIRex): Simple, single-step, and scar-free mutagenesis of bacterial genes. *PLoS One* **12**, e0184126 (2017).
43. Miller, A.W., Befort, C., Kerr, E.O. & Dunham, M.J. Design and use of multiplexed chemostat arrays. *J. Vis. Exp.*, e50262 (2013).
44. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).
45. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).