New Strategies for *In Vivo* Continuous Directed Evolution

by

Louis John Papa III

B.S. in Biochemistry, B.S. in Chemistry
University of Rochester, 2014

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN CHEMISTRY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February, 2020

Signature redacted

Signature of the Author: _____
Department of Chemistry
December 19, 2019

Signature redacted

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

Signature redacted

Accepted by: _____
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Department Committee on Graduate Students

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

Signature redacted

Alexander K. Shalek
Pfizer-Laubach Career Development Associate Professor
Department of Chemistry, 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
Associate Professor
Department of Chemistry, MIT
Thesis committee member

New Strategies for *In Vivo* Continuous Directed Evolution

by

Louis John Papa III

Submitted to the Department of Chemistry
on December 19, 2019 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in Chemistry

ABSTRACT

Continuous *in vivo* directed evolution facilitates the exploration of large biomolecule libraries at unprecedented speeds. By inserting a biomolecule of interest into a constantly mutating virus whose replicative capacity has been rendered dependent on the desired activity of that biomolecule, the once-separate mutagenesis, selection, and amplification steps of directed evolution are integrated into one simultaneous, self-sustaining process. This strategy has the potential to greatly accelerate the study of evolutionary hypotheses and the development of new biotechnologies. Unfortunately, current iterations are difficult to implement and largely restricted to *E. coli*. Furthermore, mutation rates are limited by the lack of simple mutagenesis methods that can focus mutations to desired portions of a viral genome. In this thesis, I describe the development of several new continuous *in vivo* directed evolution strategies and tools that overcome current limitations and expand the methodology to human cells. Using a generalizable adenovirus-based continuous directed evolution system, we evolved, directly in human cells, multiple variants of the tTA transcription factor that gained resistance to their small molecule inhibitor and piloted selection couples for evolving complex biomolecule activities. This system enables the continuous directed evolution of biomolecules that are important to human health and that function within complex networks that are absent in *E. coli*. Furthermore, biotechnologies developed directly in mammalian cells are more likely to have optimal function than biomolecules that are evolved in *E. coli* and then transferred to the mammalian cellular context. We also developed an *in vivo* targeted mutagenesis method that focuses mutations to a carefully defined DNA region of variable size. Using fusions of various DNA damaging enzymes and the T7 RNA polymerase, achieved high mutation rates without the usual toxicity associated with off-target mutagenesis. We expect this mutagenesis technique to be applicable across a wide variety of organisms and particularly useful for viral-based continuous evolution platforms. Finally, we are currently developing a new continuous evolution strategy for use in *E. coli* cells utilizing the lambda bacteriophage. If successful, this system would be much easier to monitor and tune than previous systems, and would expand the biomolecule cargo capacity by an order of magnitude.

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

Acknowledgements

Graduate school has been a far more incredible and fulfilling journey than I could have ever anticipated. Every day in lab has been filled with stimulating conversations, laughter, and camaraderie thanks to the amazing people I have worked with. I am forever indebted to my advisor Professor Matthew Shoulders for pushing me to achieve more than I thought possible. From our very first meeting, his infectious enthusiasm for unconventional and ambitious projects lured me in. Matt is the most sincere and even-keeled person I have ever met. Graduate school can be tumultuous, and the stability and guidance he provided was critical every step of the way. He has the uncanny ability to temper expectations when they run too high without dampening the spirit. At the same time, he can somehow rally morale and encourage me and others towards success against seemingly impossible odds. He has taught me the value of careful preparation and critical thinking, and to never become too attached to a first draft. Completely rewriting a paper or proposal always annoyed me, but the difference it made was always monumental, even if I did not want to admit it. Matt was always patiently and persistently pushing me to grow and not just settle for what was comfortable and easy; and the results were always worth it. Whatever his students need to progress as scientists, to get past a roadblock in research, or to develop our careers, Matt fights hard to get it and to solve the problem. The creative and collaborative environment he cultivated made doing research such a joy for the past six years. I could not have asked for a better advisor.

I also want to thank my committee members past and present. Professor Klibanov, thank you for always being a reliable source of earnest advice in both science and life in our annual meetings. Your 5.52 tutorial was one of the most memorable classes I have ever taken and it challenged me to think deeply about the fundamentals I thought I understood. TAing for you in 5.07 and in the 5.52 lab, where I got to meet the talented incoming first-year students each year, was a pleasure and a gift. Professor Shalek, thank you for your helpful feedback and insightful questions throughout my oral exams and for agreeing to fill in as my chair. Your hospitality at the 2015 NIH HRHR symposium was one of my fondest early memories from grad school. Professor Pentelute, thank you for stepping up on short notice to join my committee and for always being a friendly face on the fifth floor to chat with. The 5.54 Frontiers in Chemical Biology class you and Matt taught was an excellent introduction to the cutting edge of the field and really inspired me to dive into research that pushed the frontier.

To the directed evolution team, what a ride this project has been! Chet, you were a wonderful and patient mentor to me when I first joined the lab with basically zero biology experience and had to follow you around. "We are Chet." I had a blast with you in the Netherlands and New Orleans and working side-by-side for four years. Also, you may be a way better skier than me, but we established that I do not fear death on the slopes, so I think it evens out. Sam, you are one of the few who can match, and even at times exceed, my frenzied energy when discussing the project or pitching new ideas. I learn something fascinating and bizarre from you every day, and your enthusiasm for directed evolution and for outside-of-the-box ideas is inspiring. Your significant contributions began as soon as you joined and cannot be overstated. Working with you is so much fun. Mani, when you first joined the team and jumped into the frequent and high energy debates without hesitation, I knew you would fit right in. Your tenacity in this challenging

project is truly incredible and your optimism and positive energy never fail to brighten my day. When it comes to recombineering, fixing lab equipment, viral plaque assays, and the fine points of sailing, you seriously are an information sponge. Chris Moore was a sage as always and contributed heavily to the project. Duc was instrumental in the synthesis of the adenoviral protease inhibitor and was even willing to babysit me while I played with the glassware. Patreece, thank you for persevering during the most challenging early stages of the project. Alex, thank you for your contributions and your great audiobook recommendations. Caiden, thank you for your great work ethic, positive attitude, and help with cloning and recombineering. This work would also not have been possible without our collaborators Dr. Taco Uil and Professor Hoeben, who developed the error-prone adenoviral polymerase. Professor Hoeben was also an excellent host during our visit to the Netherlands and an invaluable source of advice. I also want to thank Dr. Vincent Butty for his help analyzing the sequencing data. This project required an army, and I am so grateful for the talented colleagues I interacted with and the great friends I made along the way.

Chris Moore was my partner in crime in the MutaT7 project, and one of the best mentors I could ever ask for. This project was a whirlwind with many ups and downs, and I was always amazed by Chris's ingenuity and resolve, especially in the 98 °F warm room with a feisty bioreactor. Late nights scanning plates with the music blasting and dinners at Champion's where we discussed science and life are some of my most cherished memories. Thank you for all that you taught me and for your friendship.

I also must thank the collagen folks past and present who shared the party bay with me and made day-to-day life an adventure. Madeline was an inexplicably kind and gracious bench and desk neighbor to me, a loud and not-as-organized bench and desk neighbor. Perhaps it is because we are both ardent plant enthusiasts. My new bench neighbor and fellow sailor, Kathryn, and I have reached a falcon tube sharing deal that has brought a new era of peace and harmony to the bay. Andrew was a well-spring of wisdom and some unbelievable and/or hilarious stories, and always kept spirits high. Chichi has my unending respect because of her superior euchre skills and because, like me, she has managed to silently expand to a second desk.

I also must not neglect my fellow, intrepid BL2+ virologists and all they did to help me along the way. Angela was patient and helped teach me the ways of plastic gowns, double gloves, and 10% bleach. I will never forget the day we ran the Boston marathon together in the pouring rain with Chris Richardson; it was a close race, but I decided to let her win. Emmanuel was one of the friendliest people I have ever met and was always ready with a joke and hearty laugh. He helped me set up the glorious Nikon TE200 fluorescence microscope in the BL2+ lab where the commoners could not reach it. Anna is a fellow virus enthusiast who fought valiantly alongside me in the Great Freezer Battle of 2019, a hard-won victory and day I will not soon forget. Jimin, thank you for being an active member of the fun committee with Kathryn and me, and also for being an expert virologist, sailing photographer, and absolute star on the lunch room debate team. Kenny is a vigilant safety officer, and, more importantly, a vigilant free food scout who guides the masses to many bountiful feasts. No one goes hungry in the Shoulders lab. Thank you, Kenny.

I also owe my undying gratitude to Night Crew, a loosely defined but tight-knit cadre of late night researchers, consisting mostly of postdocs and postdoc-like grad students.

Mahender was an early influence who helped set a convivial tone and who was always and still is willing to grab dinner and hang out. To Rebecca, a "fellow midwesterner," thank you for keeping us all in check and entertained with your dry wit. You are a formidable squash opponent, an excellent baker, and a wonderful labmate. Azade was the de facto leader of Night Crew, one of the coolest people I have ever known, and an important role model for me who enriched my time here more than I can express. She was the life of the party in the Shoulders lab. Thank you for continuing to be a major source of advice and support, and a dear friend. Duc is probably one of the most hilarious, generous, and thoughtful people I have ever met. He has treated me like family ever since we joined the lab together and has been a great friend who I can always count on. But he is merciless on the badminton and squash court, and destroys me every time, so it kind of cancels out.

To all the members of the Shoulders lab—postdocs, grad students, UROPs, and interns alike—thank you for the countless movie nights, verbal sparring in the break room, barbecues, and hallway chats. Working with all of you has been an incredible experience and I will cherish the memories for the rest of my life.

I must also thank the fellow members of my cohort who contributed to my journey. To Chris Richardson, who joined the Shoulders lab with me back in 2014, you have honestly made me better at arguing, which I did not think was possible. The lively discussions you sparked in lab on a daily basis, the lunch runs, the sailing voyages, and of course your friendship were invaluable to me. To the old 5.07 crew—Kellie, Shuo and Alex—thank you for making my first semester a blast. Alex, thank you for dragging me to my first group meeting at the Shoulders lab and putting them on my radar, for being a good friend, and for always being down for lunch. Shout out to Lisa and Anmol next door in the Nolan lab, who were important sources of commiseration and support, especially in the reading room (a.k.a. torture chamber) when preparing for oral exams.

Before I came to the Shoulders lab, Professor Alice Ting and all the member of the Ting lab generously welcomed me to MIT in the summer of 2014. Thank you all for your incredible hospitality and mentorship in those early days. And before that, Professor Rudi Fasan and my graduate student mentor, John Frost, at the University of Rochester made all of this possible. John and Professor Fasan dedicated an astounding amount of time and energy into training me as an undergraduate, and taught me countless lessons in chemistry, biology, writing, and public speaking. They held me to a high standard and treated me—a young, naïve undergrad—like a colleague. My time with them shaped my entire trajectory and is the reason I went into research. I also want to thank the entire Fasan lab for the fun, welcoming atmosphere that made my time as an undergraduate so wonderful. I also benefitted greatly from the mentorship of Sabri Chiha and Professor Hans-Dieter Arndt during an eventful summer 2013 internship at the University of Jena. Thank you.

And through it all, the love and support of my family carried me. Mom and Dad, you are directly responsible for fostering my interest in science from an early age. You were always in my corner and believed in me when I doubted myself. You have supported me and my dreams unconditionally, and I cannot describe how meaningful and liberating that is to me. Every day you exemplify what I can only strive to be: strong, loving, and compassionate to those around you. Without your help during all the hectic moves and your long pep talk phone calls, I seriously never would have made it to this point. Jonny

and Emily, no one truly understands me the way you do. You two are the other part of my soul and my ultimate source of refuge. You all are my rock, and I love you more than words can describe.

Table of Contents

List of Figures

Chapter 1: The Promise of Continuous *In Vivo* Directed Evolution

Chapter 2: Optimizing Recombineering for the Production of Recombinant Adenoviruses with Multiple and Large DNA Modifications

Chapter 3: Continuous Directed Evolution in Human Cells

Chapter 4: Targeted Mutagenesis by Fusing T7 RNA Polymerase to DNA Damaging Enzymes

Chapter 5: Towards a Generalizable Directed Evolution System Based on Lambda Phage Lysogeny

Chapter 6: Perspectives and Future Directions

Appendix A: β-catenin Directed Evolution

List of Tables

Chapter 2: Optimizing Recombineering for the Production of Recombinant Adenoviruses with Multiple and Large DNA Modifications

Chapter 3: Continuous Directed Evolution in Human Cells

Chapter 4: Targeted Mutagenesis by Fusing T7 RNA Polymerase to DNA Damaging Enzymes

Chapter 5: Towards a Generalizable Directed Evolution System Based on Lambda Phage Lysogeny

Appendix A: β-catenin Directed Evolution

Abbreviations

°C	Degrees Celsius
× *g*	Times gravity
µg	Micrograms
µL	Microliters
µM	Micromolar
A	Alanine or adenosine
aaRS	Aminoacyl-tRNA synthetase
AdBAC	Adenoviral bacterial artificial chromosome
ADP	Adenoviral death protein
AdPol	Adenoviral DNA polymerase
AdProt	Adenoviral protease
amp^R	Ampicillin resistance gene
attR	Right attachment site of the lambda prophage
BAC	Bacterial artificial chromosome
BOI	Biomolecule of interest
bp	Base pairs
C	Cysteine or cytidine
Cas9	CRISPR associated protein 9
ccdA	Control of cell death protein A
ccdB	Control of cell death protein B
cDNA	Complementary DNA
CFP	Cyan fluorescent protein
CFU	Colony forming units
chlor^R	Chloramphenicol resistance
CHO	Chinese hamster ovary cells
cI	Lambda phage repressor protein cI
cIG112E	A G112E mutant of cI
cm	Centimeters
CMV	Cytomegalovirus immediate-early promoter/enhancer
cos	Cohesive site of cleavage and packaging in lambda phage
CPE	Cytopathic effect
d	days
D	Aspartate
DAPI	4′,6-Diamidino-2-phenylindole
DBP	Adenoviral DNA binding protein
ddH_20	Double distilled water
DMEM	Dulbecco's modified Eagle medium
DNA	Deoxyribonucleic acid
dox	Doxycycline
drApo1	Deactivated *Rattus norvegicus* APOBEC-1
E	Glutamate
E. coli	*Escherichia coli*
E1	Early region 1 of adenovirus
E3	Early region 3 of adenovirus

E4	Early region 4 of adenovirus
eGFP	Enhanced green fluorescent protein
EP-Pol	Error-prone adenoviral DNA polymerase
F	Phenylalanine
FACS	Fluorescence-activated cell sorting
folA	*E. coli* dihydrofolate reductase gene
G	Glycine or guanosine
galK	*E. coli* galactokinase gene
GFP	Green fluorescent protein
GGS	Glycine-serine linker
Gly-Ser	Glycine-serine linker
GOI	Gene of interest
gRNA	Guide RNA
h	Hours
H	Histidine
HEK293A	Human embryonic kidney 293 cells
HIV	Human immunodeficiency virus
HSP90	Heat shock protein 90
I	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Lysine
kan^R	Kanamycin resistance gene
kb	Kilobase
kV	Kilovolt
L	Leucine
lacI	Lac operon repressor gene
lacZ	*E. coli* β-galactosidase gene
lacZα	*E. coli* β-galactosidase alpha fragment gene
lacZω	*E. coli* β-galactosidase omega fragment gene
LB	Lysogeny broth, or Luria-Bertani broth
LED	Light-emitting diode
LeuRS	*E. coli* leucyl-tRNA synthetase
Lsel	Helpher lambda prophage lysogen strain
M	Methionine or molar
mg	Milligram
min	Minute
minCMV	Minimal CMV promoter
mL	Milliliter
mM	Millimolar
MOI	Multiplicity of infection
MP6	Mutagensis plasmid 6
msec	Millisecond
MutaT7	DNA damaging enzyme–T7 RNA polymerase fusion protein
MW	Molecular weight
N	Asparagine or sample size
n	Sample size

N.D.	Not detected
ng	Nanogram
nM	Nanomolar
NRPS	Non-ribosomal peptide synthetase
OD	Optical density
oligo	Oligonucleotide
*o*NPG	*ortho*-Nitrophenyl-β-galactoside
OPTI-MEM	Reduced serum derivative of minimal essential medium
P	Proline
$P_{A1LacO-Tenth}$	A weak lacI-repressible promoter
PACE	Phage-assisted continuous evolution
P_{BAD}	Arabinose-inducible promoter
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEI	Polyethyleneimine
PKS	polyketide synthases
P_L	Leftward promoter of lambda phage
P_R	Rightward promoter of lambda phage
P_{T7}	T7 promoter
P_{tac}	Constitutive tac promoter
pTP	Adenoviral preterminal protein
PVI	Adenovirus precursor protein VI
PylRS	Pyrrolysyl-tRNA synthetases
Q	Glutamine
qPCR	Quantitative PCR
R	Arginine
r.p.m.	Rotations per minute
rApo1	*Rattus norvegicus* APOBEC1
recA	*E. coli* recA recombination protein
recA1202	A recA mutant with hyperactive co-protease activity
REV	HIV protein rev
RNA	Ribonucleic acid
rpsL	*E. coli* 30S ribosomal protein S12 gene
RRE	Rev response element
RT-qPCR	Quantitative reverse transcriptase qPCR
rtTA	Reverse tetracycline transactivator
S	Serine
S.D.	Standard deviation
s.e.m.	Standard error of the mean
scFv	Single-chain variable fragment
SM	Lambda phage suspension medium
SOC	super optimal broth with catabolite repression
stf	Lambda phage side tail fiber gene
$Strep^R$	Streptomycin resistance phenotype
Str^R	Streptomycin resistance phenotype
T	Threonine or thymidine

tadA	Adenosine deaminase
TALEN	Transcription activator-like effector nuclease
TAR	Trans-activating response element
tat	HIV trans-activating regulatory protein
tBcat	TetR–β-catenin fusion protein
tetO	Tetracycline responsive element
tet^R	Tetracycline resistance phenotype
TetR	Tetracycline repressor
TMP	Trimethoprim
TP	Adenoviral terminal protein
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
tRNA	Transfer RNA
tTA	Tetracycline transactivator
tTA_{mut}	Mutant tetracycline transactivator that binds
tTA_{WT}	Wild-type tetracycline transactivator
TTC	Tetrazolium chloride
ugi	Uracil DNA glycosylase inhibitor
ung	Uracil DNA glycosylase gene
UV	Ultraviolet
V	Valine
VSVG	Vesicular stomatitis virus envelope glycoprotein
W	Tryptophan
X-gal	5'-bromo-4-chloro-3-indolyl-B-galactopyranoside
Y	Tyrosine
Δ	Genetic deletion
σ^{70}	*E. coli* sigma 70 transcription factor

Chapter 1: The Promise of Continuous *In Vivo* Directed Evolution

1.1 Overview

Nature is teeming with spectacular solutions to the innumerable challenges organisms face across diverse environments. From rapid cheetahs that can catch agile antelopes, to hardy Greenland sharks that can survive frigid waters, to insects that perfectly resemble leaves and sticks, it is readily apparent that the difficult circumstances that abound in an ever-changing world—such as faster prey, colder seas, or increasingly sharp-eyed predators—are merely opportunities for creatures with amazing attributes to arise and thrive with little competition. The process of natural evolution gradually sculpts such exquisite adaptations over time through random variation and selection.

Random variation by mutation generates a spectrum of traits, selection pressure in the environment eliminates individuals with less fit traits on the spectrum and liberates resources for individuals with more fit traits, and the resulting expansion of populations with beneficial traits gives rise to a new spectrum of traits and increases the probability of even better traits appearing. For example, if a large population of slow, prehistoric cheetahs encounters faster prey on the savanna, only the few cheetahs that happen to be faster than the antelope are likely to survive. These few, fast cheetahs will thrive on the abundant, previously-uncatchable antelope to produce a new generation of cheetahs that inherit the speed of their parents. Some of the new offspring may be even faster than their parents owing to random variation, with a large population making such a rare mutant cheetah more likely to occur. As generations proceed, the average speed of the cheetah population gradually increases until the modern population is remarkably faster than the original prehistoric cheetahs, thus adding to the pantheon of evolutionary masterpieces.

The principles of this cycle are evident both in the macroscopic world of animals and in the microscopic world of microbes and individual cells, where evolution often has important implications for human health. Evolution gives rise to antibiotic-resistant bacteria,[1] cancer cells that proliferate unchecked and metastasize,[2] and viruses that evade host immune responses.[3-8] Studying the nuances of mutation rates, selection pressures, and evolutionary trajectories is important for understanding, preventing, and treating many human diseases. However, the environments and conditions that result in the evolution of disease or new species need not only be passively observed in nature. By actively designing and creating artificial environments in the laboratory, one can mimic and accelerate the trial-and-error process of natural evolution to test fundamental evolutionary hypotheses, and to generate new biomolecules with useful, novel functions as tools and therapies.

Unlike purely rational attempts to engineer biomolecules, this “directed evolution” approach does not require deep, prior structural or mechanistic knowledge of the parent biomolecule to succeed. The enormous impact of directed evolution is reflected in the staggering swath of tailored biomolecules and tools it has produced, including industrial enzymes,[9,10] amino-acyl tRNA synthetases for unnatural amino acid incorporation,[11,12] transcription factors,[13,14] proteases,[15] repressors,[16] recombinases,[17,18] inteins,[19] membrane transporters,[20] targeted DNA binders,[21,22] G protein-coupled receptors,[23] DNA base editors,[24,25] RNA polymerases, [26,27] antibodies,[28,29] fluorescent proteins,[30] and much more. This impact was also acknowledged in the awarding of the 2018 Nobel Prize in Chemistry to Frances Arnold for her work in the directed evolution of enzymes, and to George Smith and Sir Gregory Winter for their work in directed evolution through phage display.[31]

Most directed evolution strategies are performed partially or entirely *in vitro*, where mutagenesis, selection, and amplification are performed as separate steps (**Figure 1.1**).[9,32] Typically, these approaches entail generating a library of mutant genes *in vitro* through techniques like error-prone PCR, which is then transformed into *E. coli* or yeast for expression. In some cases, the desired function of the biomolecule of interest (BOI) is selected by somehow coupling it to the survival of the BOI-expressing organism, such as by linking BOI activity to the expression of an antibiotic resistance gene. In other cases, the desired activity is screened for by assaying BOI-expressing organisms or purified BOI variants individually. Once one or more BOIs with improved activity are found, the DNA encoding these BOIs is isolated, sequenced, and mutated further for the next round of selection or screening. While the traditional discontinuous approach has yielded many useful, new biomolecules, each step requires significant manipulation and time to perform, thus constraining the number of variants that can be explored with reasonable effort.

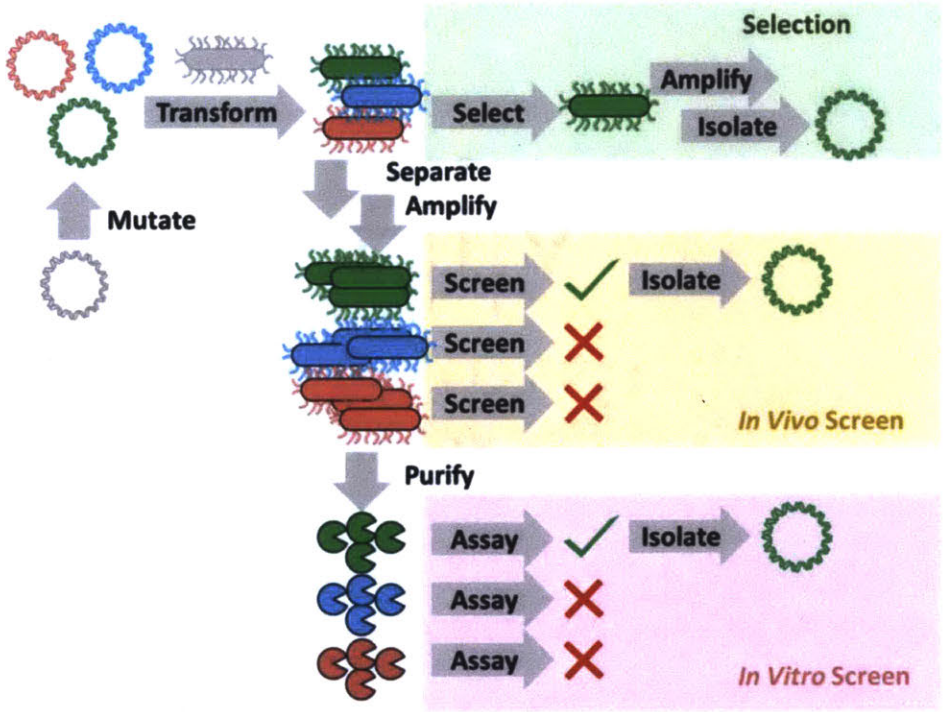

Figure 1.1 Typical directed evolution workflows.
A schematic of the required steps in a variety of traditional directed evolution workflows. Mutagenesis is typically performed *in vitro* and DNA plasmid libraries are then transformed into yeast or *E. coli*. If the desired BOI activity is important for organism survival, it can be selected for (green box). Otherwise the activities of individual BOI expressing organisms need to be screened *in vivo* (yellow box), or the activities of purified biomolecules need to be screened *in vitro* (red box).

Aside from low-throughput, another major limitation of traditional directed evolution approaches is that BOI activities are often evolved in an alien context. BOIs evolved purely *in vitro* or in heterologous organisms, such as yeast or *E. coli*, often fail to perform the desired function when transported into the more complex organisms in which they

were intended to function, such as human cells. There are several published examples of evolved intrabodies,[33] scFvs,[34,35] inteins,[19] and cellular receptors[23] selected *in vitro* or in lower organisms that had reduced or no activity in human cells. Moreover, most examples of BOIs that fail to translate from lower organisms into human cells remain unpublished. Precise reasons for failure are rarely elucidated, but evolved protein functions are likely derailed in complex mammalian cells by unintended intermolecular interactions, poor folding, unexpected cellular localization or post-translational modifications, and many other serious problems. Furthermore, certain classes of biomolecules function as components of complex systems or networks that are entirely absent in lower organisms, and therefore their activities can only be selected or evolved within their native context.

Some of the major limitations of conventional directed evolution approaches can be mitigated with alternate strategies that more closely mimic natural evolution. Recently developed *in vivo* continuous evolution systems, such as phage-assisted continuous evolution (PACE)[26]—in which the mutation, selection, and amplification steps occur constantly and simultaneously in a reproducing population—have greatly accelerated the directed evolution cycle. Shorter cycle times and less labor per cycle allow many more variants to be screened and thus increase the chance of encountering rare, improved variants.[32] Furthermore, continuous *in vivo* evolution of BOIs directly within their intended context guarantees optimal activity. This principle was strikingly illustrated by the successful continuous directed evolution of a human-optimized rtTA variant directly within human cells by coupling HIV replication to the activity of an HIV-borne rtTA gene.[13,36] While continuous *in vivo* systems show great promise, they are not widely adopted owing to unique challenges to their implementation that are not as pronounced in conventional discontinuous approaches.[32] For example:

1) Tunable genetic circuits must be designed to tightly link the desired activity to organism survival.
2) The opportunity for false positives to proliferate must be reduced or eliminated.
3) Extant *in vivo* mutagenesis methods are largely untargeted and inferior to *in vitro* methods.

Eliminating these trade-offs would foster widespread application of continuous *in vivo* directed evolution and accelerate efforts to develop new therapeutics and biomolecular tools. Below, I review each of these unique challenges in detail.

1.2 The importance of selection circuit design and optimization

When employing discontinuous evolution, the desired activity need not affect organism survival and can rather be screened directly *in vitro* or within the BOI-expressing organism. For example, the activities of enzymes that perform chemical reactions with small molecules can be quantified using analytical methods such as liquid or gas chromatography and/or mass spectrometry,[37] and the activities of fluorescent proteins can be screened using fluorescence activated cells sorting (FACS).[30] Nonetheless, the limited capacity to analyze each and every variant constrains the size of the library that can realistically be screened. To achieve continuous *in vivo* evolution and circumvent such laborious analyses, BOI activities that might not inherently impact organism fitness must be tightly coupled to host survival using a specially-designed genetic selection circuit. The most straightforward selection circuit is one in which the desired BOI activity induces expression of a critical or conditionally critical gene. For example, to evolve a transcription factor that binds a new DNA sequence an antibiotic resistance gene can be placed downstream from the new DNA sequence. Transcription factor variants that bind the new sequence and express the antibiotic resistance gene will allow the host cell carrying that particular variant to survive in the presence of an antibiotic while neighboring cells carrying non-functional variants will be killed by the antibiotic.

There are many clever examples in the literature for how to couple the expression of a critical gene to the activity of various BOIs,[12,15,21,25-27,38] such as critical genes with premature stop codons for evolving amino-acyl tRNA synthetases that charge stop codon suppressor tRNAs,[11,12] or expression of a critical gene from a two-hybrid system in order to evolve novel protein binders.[38] Indeed, selections in which organism survival depends on BOI activity have facilitated the exploration of large libraries in both continuous and discontinuous formats. However, careful optimization of selection pressure is critical to success. If selection pressure is too low after a discrete selection step, organisms carrying low activity variants will still survive and outnumber the rare individuals carrying improved variants. If selection pressure is too low during continuous selection, improved variants will only confer a minimal advantage to their host and will require many generations to enrich in the population. Conversely, if selection pressure is too high, the probability of any variants appearing with sufficient activity for survival is very low and the entire population dies out, leaving no variants at all.

Essentially, selection pressure must always be appropriate and typically requires gradual adjustment after each round of evolution, when the average activity of the library increases and as higher activities are sought. In a discontinuous format, the results of a selection are manually observed after each round and the round must be repeated with altered selection pressure if that round failed to produce improved variants. Alternatively, in a continuous evolution system, selection pressure must be constantly balanced such that the population is just barely surviving over a long period of time, but never dies out. If the population of BOI-expressing organisms begins to diminish, the selection circuit must be tuned to allow higher expression of the critical gene with lower BOI activity, thus allowing population recovery before it dies out. If the average BOI activity of the population improves to the point that it induces superfluous levels of critical gene expression and there is no selection pressure, the selection circuit must be tuned to require higher BOI activity for lower critical gene expression, thus maintaining sufficient selection pressure for the enrichment of new variants. In an ideal continuous evolution system, it must be

easy to monitor the size of the population. It also must be possible to adjust the selection pressure accordingly.

The PACE selection circuit functions by coupling the desired BOI activity to the expression of the pIII coat protein that is required for efficient M13 bacteriophage infection (**Figure 1.2**). The BOI-inducible pIII gene is located on a plasmid within the *E. coli* host, and the genes for the BOI variants are located in a population of propagating M13 phage. If a cell is infected with a BOI variant that has sufficient activity, that cell will produce infectious, pIII-containing M13 capsids carrying that variant that will go on to infect more cells. Cells infected with low or no activity variants produce non-infectious capsids, and that variant eventually disappears as the M13 phage population continually reproduces to overcome constant dilution. Careful selection pressure optimization has been critical to the success PACE has had in producing a wide variety of new biomolecules. PACE selection pressure modulation has been achieved by altering the copy-number or expression strength of the selection circuit, supplementing pIII expression from a small molecule-inducible promoter, increasing the rate of population dilution, and/or creating different "stepping stone" intermediate substrates that are a hybrid between the parent and target substrates of the desired activity.[15,21,22,25-27,38-43]

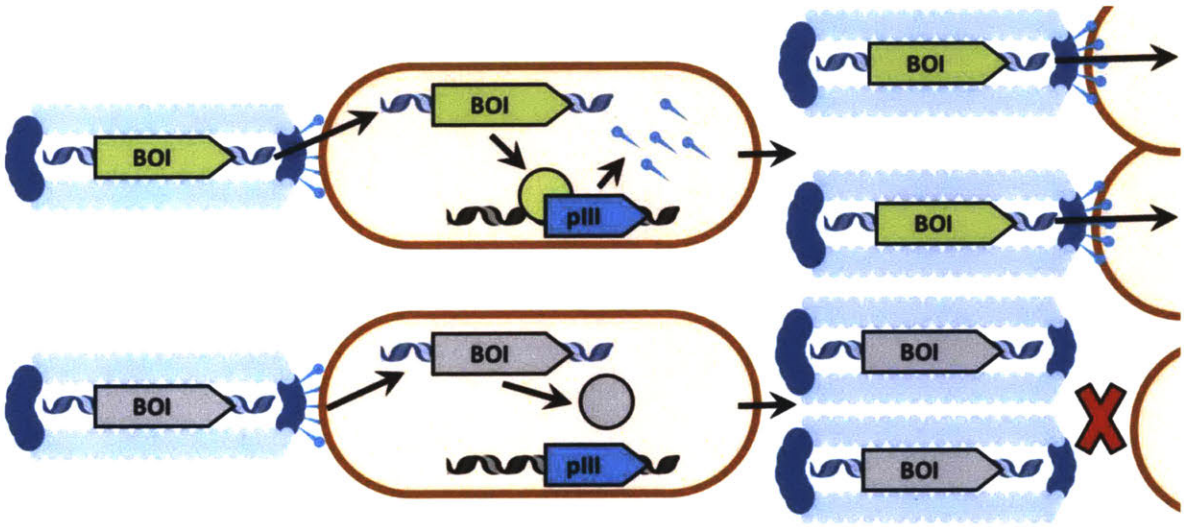

Figure 1.2 The PACE selection circuit.
Active BOI variants (green) are able to induce pIII expression from the selection circuit and thus produce infectious M13 phage carrying that active variant. Inactive BOI variants (gray) only yield M13 phage that are unable to infect new cells efficiently.

Nonetheless, selection pressure optimization in PACE is complicated by the inability to directly measure the size of the M13 phage population in real time during evolution. The M13 phage is submicroscopic and does not kill its host,[44] unlike other phage species that lyse their host cells and cause the turbid culture to turn clear. In a cloudy culture of bacteria, which essentially serve as the "food" for the M13 population in PACE, there is no immediate way to determine whether the M13 phage population is thriving or has gone completely extinct. If a lytic phage were used, the bacteria would lyse and the culture would become less cloudy, providing a simple optical cue as to whether new progeny phage are being produced or not. If bacteria themselves carried the BOI and served as the propagating "population" rather than the "food", the cloudiness of the culture is directly correlated with population size. In the case of a lytic phage or bacterial

continuous evolution, population size and thus selection pressure could automatically be regulated with a standard turbidostat. Instead, in PACE, the titer of the phage population is usually determined by taking a small sample of the evolving population and using a labor-intensive plaque assay that usually must develop overnight. This ~16 hour delay in evolution monitoring requires one to use conservative selection pressure, such as slow dilution rates, or accept the risk of population extinction and restarting the experiment. Attempts have been made to infer average BOI activity in real time by coupling both pIII and luciferase expression to BOI activity and measuring luciferase activity with a luminescence detector built into the bioreactor setup.[27] However, this approach requires specialized equipment and has not been widely applied to monitor activity in real time. Rather, standard luminescence plate assays have largely been used to pre-optimize PACE selection conditions and to characterize variants that arise from the bioreactor.[15,21,22,25,26,38-43] Thus, PACE is able to undergo many evolution cycles at unprecedented speed, but ultimately requires specialized equipment and expertise.[32]

In the HIV-based metazoan system, the transcription of HIV genes was placed under the control of the exogenous rtTA transcription factor, which was inserted into the HIV genome (**Figure 1.3**).[45,46] Normally HIV transcription is controlled by HIV tat protein binding to the TAR hairpin, but by replacing tat with rtTA and TAR with the tetO operator, HIV replication was tightly linked to rtTA activity instead.[46] This virus was then passaged for many generations to evolve rtTA to be more active and more sensitive to its small molecule activator, doxycycline.[13,36] Selection pressure was applied by periodically passaging and thus diluting the virus, favoring variants that can replicate fast enough to replace the viruses that were lost and thereby outcompete neighboring variants. Selection pressure was balanced by only passaging the virus once the viral population had reached peak production and was forming wide-spread, visible multinucleated cells or "syncytia," allowing enough time for the population to recover after each bottleneck.[36,45] While both this HIV-based system and PACE have varied the timing or rate, respectively, of population dilution as a means of selection pressure modulation, the HIV-based system had the benefit of being easily monitorable in the microscope through the formation of obvious, tell-tale syncytia, thus allowing the prevention of a complete die off of the population.

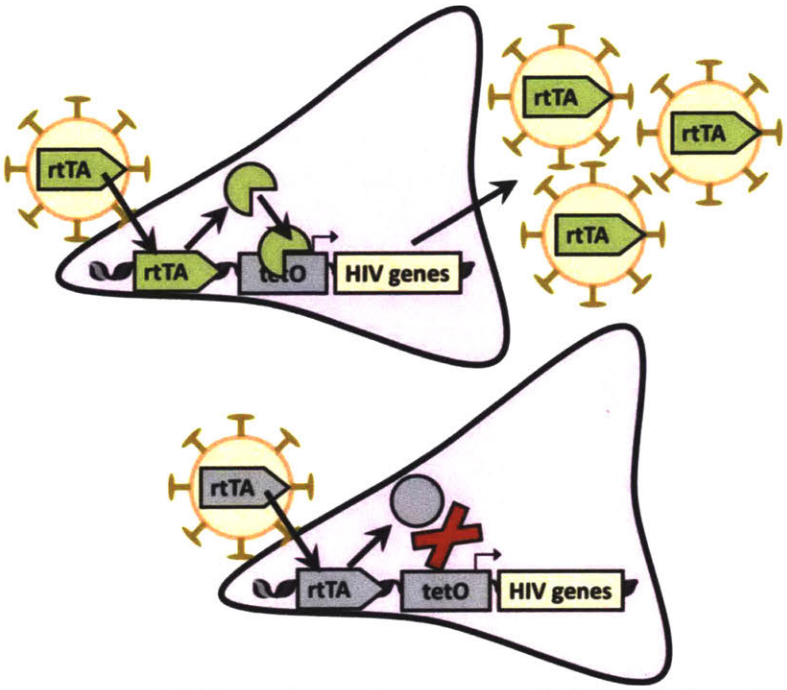

Figure 1.3 HIV-based continuous evolution of the rtTA transcription factor.
Active BOI variants (green) are able to induce expression of the HIV genome from the selection circuit and thus produce HIV virus carrying that active variant. Inactive BOI variants (gray) are unable to induce the HIV genes and no virus is formed.

While the HIV-based system was drastically slower than PACE, it expands the continuous evolution toolbox by evolving BOIs specifically in human cells, which may be critical for certain target BOIs. One could envision generalizing the HIV-based system for the evolution of other biomolecules by coupling rtTA expression to various BOIs that can be inserted into the HIV genome. However, this platform was not explored further in the literature and it would likely not be widely adopted owing to the biosafety hazards associated with large-scale HIV culture. New selection systems that are safe, available in a variety of host organisms, easily tunable, and easy to monitor in real time are still needed to fully realize the potential of continuous *in vivo* directed evolution.

1.3 Prevention of "cheaters"

During directed evolution, false positives or "cheaters" are individuals in the population that have evolved a mechanism other than the desired activity to overcome selection pressure. For example, if attempting to evolve a transcription factor to recognize a new promoter and turn on an antibiotic resistance gene, an *E. coli* cell with an inactive variant of the transcription factor can still cheat the selection by acquiring mutations upstream of the antibiotic resistance gene that turn it on in the absence of transcription factor activity. Other mechanisms of cheating include the advantageous mutation of genes outside of the selection circuit, such as mutation of native drug efflux pumps to overcome antibiotic selection. Cheaters often have a significant advantage over all variants, because they have somehow lost BOI dependence, and will gradually outcompete all other variants in the population, including true positives, if they are not prevented or eliminated.

In a traditional discontinuous selection format, most cheaters are eliminated at the end of each round when the DNA of the BOI plasmid library is isolated and then retransformed into a fresh isogenic host for the next round. Unlike mutations in the BOI library, mutations in the selection circuit or host genome are not carried over from round to round, so cheaters rarely get the chance to outnumber true positives. Discontinuous evolution automatically eliminates many potential cheaters by virtue of DNA isolation for characterization, further mutation, and retransformation. In contrast, the prevention of cheaters needs to be consciously factored into the design of any continuous directed evolution system and is not as straightforward to accomplish. Since there are no organisms that can perfectly replicate polynucleotides indefinitely without any errors, cheaters that mutate the selection circuit or native host genes are bound to arise eventually. It is thus practically a necessity to constantly replace the selection circuit during long-term continuous directed evolution.

The most practical way to achieve selection circuit replenishment in a continuous manner is to use a phage or a virus to encode the BOI, and to place the selection circuit in a constantly replenished host. When new phage or virus exits the host, it is akin to DNA purification of the BOI library in the discontinuous format; and when a phage or virus infects a new host, it is akin to transformation. Thus selection circuits that mutate to be BOI-independent are not allowed to propagate, as they are lysed by the phage or virus or lost to constant dilution or passaging, and only the virus or phage is allowed to propagate. An added benefit is that virus and phage genomes encode drastically fewer genes than cell genomes, and thus generally have fewer potential cheating mechanisms than an entire evolving cell.[47] PACE implemented this strategy by making M13 bacteriophage dependent on an external accessory plasmid that conditionally expressed pIII and was not allowed to propagate.[26] The HIV-based system was not comparable in that the entire selection couple (i.e. rtTA, the tetO operator, and the HIV genes) was encoded in the propagating HIV genome. It is surprising that the HIV genome did not evolve to become rtTA independent, especially given that tat and TAR were inactivated by only a handful of point mutations.[46] It is therefore not clear if the rtTA strategy would be viable for the evolution of other BOI activities. Future viral-based strategies in human cells should ideally separate the genetic circuit from the viral genome to prevent cheater propagation.

1.4 Challenges of *in vivo* mutagenesis

A major benefit of traditional discontinuous evolution is the plethora of sophisticated tools for performing mutagenesis on DNA plasmids *in vitro*. Diverse libraries can easily be generated with highly mutagenic error-prone PCR and DNA shuffling. Mutations can also be hyper-focused using site-directed mutagenesis to create double, triple, or more mutations only in regions of interest.[48] Unfortunately, tools for introducing mutations in living cells that do not require DNA isolation are far less developed, and generally rely on high energy radiation, chemical mutagens,[49,50] or expressing mutagenic enzymes.[51,52] These methods introduce mutations globally throughout the entire genome and not just to the gene of interest. Since the gene of interest typically only constitutes a tiny fraction of the DNA present in a cell or a viral genome, a high mutagenesis rate is required to obtain any significant number of mutants. It follows that most mutations will be "off-target". Many of these off-target mutations will occur in essential genes and kill the cells[53,54] or viruses[55] in which they occur, requiring a balance between library diversity and mutagenic toxicity that constrains library size and limits the ability to explore double, triple, or quadruple mutants. If the mutation rate is very high, more mutations will occur in the gene of interest, but the probability of simultaneous lethal mutations occurring increases, thus masking these mutants. If the mutation rate is very low, most individuals will survive the mutation process but library diversity will be low, making it unlikely that an improved variant will appear (**Figure 1.4**).

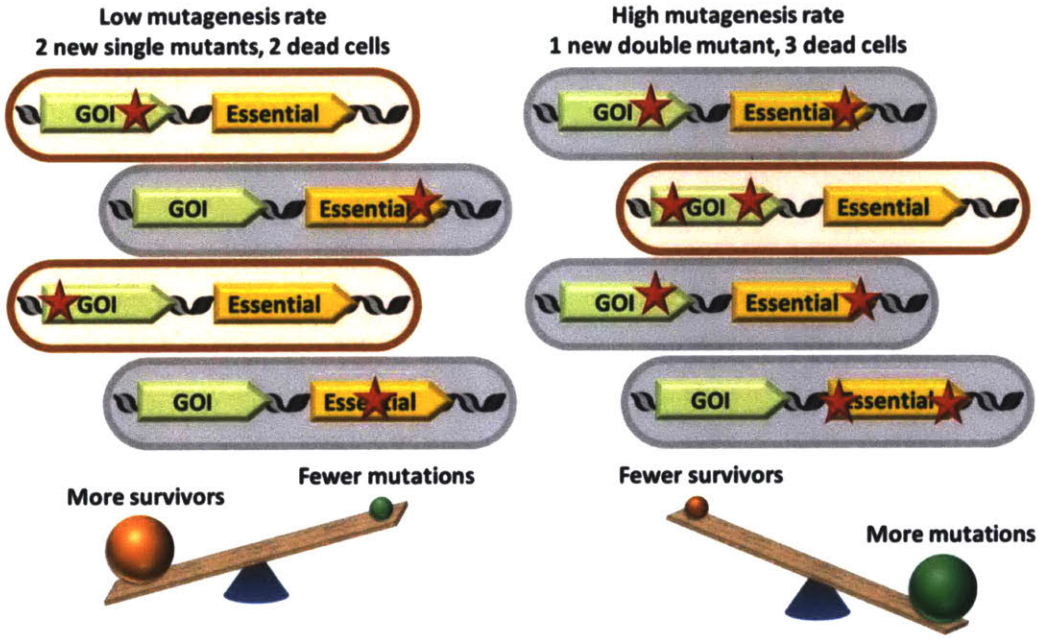

Figure 1.4 Global mutagenesis rates must balance diversity and toxicity.
Hypothetical population mutation scenarios are shown. In a low mutagenesis rate scenario (left), only two new variants appear in the gene of interest (GOI). In a high mutagenesis rate scenario (right), three new GOI variants appear, two of which are removed from the population owing to simultaneous lethal mutations in genes essential for organism survival.

PACE relies on the expression of globally mutagenic genes from an inducible plasmid.[52] The HIV-based system relies on the high intrinsic error rate of HIV reverse transcription.[13] Both systems suffer from the error catastrophe limit inherent to global mutagenesis. If mutations were targeted only to the gene of interest in these systems and

not to the rest of the phage or viral genome, very high mutation rates could be achieved and increase the chance of producing an improved variant without killing the phage or virus. Several targeted *in vivo* mutagenesis systems have been recently developed, but none are ideal for a continuous phage- or viral-based evolution system. Fusions between Cas9 and DNA damaging enzymes or error-prone polI can edit bases within a narrow window,[24,56-60] but many guide RNAs would have to be designed, synthesized and expressed to cover an entire gene that is more than a few hundred base pairs.[61,62] Critically, guide RNAs would also likely lose effectiveness in important regions as mutations accumulate, requiring the identification of new mutations and redesigning of guide RNAs, which is not practical in a continuous format. In another method, mutations can be targeted near a ColE1 origin using an error-prone polI variant, but this is specific only to ColE1 plasmids in *E. coli*[63,64] and likely cannot be applied to M13 phage or mammalian viruses. Error-prone orthogonal replication systems also exist for yeast using Ty1 retrotransposons[65] or yeast plasmids,[66] but these are limited to yeast and are not applicable to a phage- or virus-based system.

In sum, current *in vivo* targeted mutagenesis methods are not well-suited for phage- or viral-based continuous evolution because they are either limited to very specific genetic elements or have a narrow editing window that requires target specific guide RNAs. An ideal targeted mutagenesis system should be able to act over a long period of time without the need for guide RNA redesign, and should act upon genes of any size in a wide variety of contexts.

1.5 Summary

While PACE provides a rapid approach for evolving biomolecules, it can only do so in the context of *E. coli* cells. Furthermore, it is difficult to implement owing, at least in part, to the lack of real time feedback on phage population size and the resulting risk of total population die off. HIV-based directed evolution is easily monitored and has been used to successfully evolve human-optimized rtTA directly in human cells. However, the HIV-based system has not been developed further and would likely pose a serious biosafety hazard. Additionally, the rtTA based selection couple is not separated from the HIV genome and would likely be prone to mutation accumulation and loss of BOI-dependence, giving rise to cheaters. Finally, the mutation rates of both PACE and HIV-based evolution are currently limited by global mutagenesis toxicity and there are no targeted *in vivo* mutagenesis tools that can easily be applied specifically in the context of phage- or viral-based continuous directed evolution.

In Chapters 2 and 3, I describe the development of a generalizable, safe adenovirus-based continuous directed evolution system in human cells. Chapter 2, outlines the optimization and employment of lambda *red* recombineering for precision editing of adenoviral genomes to construct the necessary vectors. Chapter 3 describes how the resulting viruses were used to evolve multiple doxycycline-resistant variants of the tTA transcription factor directly in human cells. We also demonstrated viable selection couples for the evolution of site-specific recombinases and aminoacyl-tRNA synthetases. We believe this system will serve as a powerful tool for studying disease-relevant evolutionary trajectories directly in human cells and for developing biotechnological tools and therapies that are virtually guaranteed to work within a human cellular context.

In Chapter 4, I describe the development of a targeted mutagenesis method that can target large, but carefully defined, DNA regions. By fusing DNA damaging enzymes to the T7 RNA polymerase, we were able to achieve very high mutation rates selectively between a T7 promoter and a T7 terminator array without toxicity. We believe that this system is particularly useful for phage- or viral-based continuous directed evolution, and should be applicable across various organisms from *E. coli* to human cells.

In Chapter 5, I describe preliminary results toward the development of a phage-based continuous evolution system in *E. coli* that is easy to implement and monitor in real time. The selection couple is based off the well-characterized lysogeny-lytic genetic switch of lambda phage. By coupling BOI activity to the propagation of a BOI-carrying cosmid in specially-engineered helper lysogen cells, we can monitor the progress of evolution through simple optical monitoring of host cell lysis. Preliminary data also suggest that selection pressure can easily be tuned by temperature using the temperature-sensitive cI857 repressor variant. This system would have a BOI capacity of 53 kb,[67] compared to the ~2 kb capacity of M13 bacteriophage,[68] because all lambda phage genes are absent from the phage capsid, which only contains the BOI cosmid. Also, very high BOI mutation rates should be achievable because mutations only accumulate in the BOI cosmid and not in the lambda phage genes. We believe this system would offer a simple, user-friendly alternative to PACE.

In Chapter 6, I provide my perspective on the current state of the field and future outlooks. The appendix reviews efforts to evolve the oncogenic β-catenin transcription factor, which functions within the context of a complex cellular network that is absent in *E. coli* and yeast. Initial efforts were unsuccessful, but provided several useful lessons

that could lead to an improved, highly optimized version of our human cell-based continuous directed evolution platform.

1.6 References

[1] Davies, J., and Davies, D. (2010) Origins and evolution of antibiotic resistance, *Microbiol. Mol. Biol. Rev. 74*, 417–433.

[2] Greaves, M., and Maley, C. C. (2012) Clonal evolution in cancer, *Nature 481*, 306–313.

[3] Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M., Saag, M. S., Komarova, N. L., Nowak, M. A., Hahn, B. H., Kwong, P. D., and Shaw, G. M. (2003) Antibody neutralization and escape by HIV-1, *Nature 422*, 307–312.

[4] Rambaut, A., Posada, D., Crandall, K. A., and Holmes, E. C. (2004) The causes and consequences of HIV evolution, *Nat. Rev. Genet. 5*, 52–61.

[5] Carrat, F., and Flahault, A. (2007) Influenza vaccine: the challenge of antigenic drift, *Vaccine 25*, 6852–6862.

[6] Phillips, A. M., Doud, M. B., Gonzalez, L. O., Butty, V. L., Lin, Y. S., Bloom, J. D., and Shoulders, M. D. (2018) Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin, *eLife 7*, e38795.

[7] Phillips, A. M., Gonzalez, L. O., Nekongo, E. E., Ponomarenko, A. I., McHugh, S. M., Butty, V. L., Levine, S. S., Lin, Y. S., Mirny, L. A., and Shoulders, M. D. (2017) Host proteostasis modulates influenza evolution, *eLife 6*, e28652.

[8] Phillips, A. M., Ponomarenko, A. I., Chen, K., Ashenberg, O., Miao, J., McHugh, S. M., Butty, V. L., Whittaker, C. A., Moore, C. L., Bloom, J. D., Lin, Y. S., and Shoulders, M. D. (2018) Destabilized adaptive influenza variants critical for innate immune system escape are potentiated by host chaperones, *PLOS Biol. 16*, e3000008.

[9] Jäckel, C., Kast, P., and Hilvert, D. (2008) Protein design by directed evolution, *Annu. Rev. Biophys. 37*, 153–173.

[10] Cherry, J. R., and Fidantsef, A. L. (2003) Directed evolution of industrial enzymes: an update, *Curr. Opin. Biotechnol. 14*, 438–443.

[11] Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of Escherichia coli, *Science 292*, 498–500.

[12] Sharma, V., Zeng, Y., Wang, W. W., Qiao, Y., Kurra, Y., and Liu, W. R. (2018) Evolving the N-terminal domain of pyrrolysyl-tRNA synthetase for improved incorporation of noncanonical amino acids, *ChemBioChem 19*, 26–30.

[13] Das, A. T., Zhou, X., Vink, M., Klaver, B., Verhoef, K., Marzio, G., and Berkhout, B. (2004) Viral evolution as a tool to improve the tetracycline-regulated gene expression system, *J. Biol. Chem. 279*, 18776–18782.

[14] McKeown, A. N., Bridgham, J. T., Anderson, D. W., Murphy, M. N., Ortlund, E. A., and Thornton, J. W. (2014) Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module, *Cell 159*, 58–68.

[15] Dickinson, B. C., Packer, M. S., Badran, A. H., and Liu, D. R. (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations, *Nat. Commun. 5*, 5352.

[16] Ellefson, J. W., Ledbetter, M. P., and Ellington, A. D. (2018) Directed evolution of a synthetic phylogeny of programmable Trp repressors, *Nat. Chem. Biol. 14*, 361–367.

[17] Lansing, F., Paszkowski-Rogacz, M., Schmitt, L. T., Martin Schneider, P., Rojo Romanos, T., Sonntag, J., and Buchholz, F. (2019) A heterodimer of evolved

designer-recombinases precisely excises a human genomic DNA locus, *Nucleic Acids Res.*, gkz1078.
[18] Meinke, G., Karpinski, J., Buchholz, F., and Bohm, A. (2017) Crystal structure of an engineered, HIV-specific recombinase for removal of integrated proviral DNA, *Nucleic Acids Res. 45*, 9726–9740.
[19] Peck, S. H., Chen, I., and Liu, D. R. (2011) Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells, *Chem. Biol. 18*, 619–630.
[20] Bali, A. P., Genee, H. J., and Sommer, M. O. A. (2018) Directed evolution of membrane transport using synthetic selections, *ACS Synth. Biol. 7*, 789–793.
[21] Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., Tsai, S. Q., Sander, J. D., Joung, J. K., and Liu, D. R. (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity, *Nat. Methods 12*, 939–942.
[22] Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, *Nature 556*, 57–63.
[23] Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S., and Roth, B. L. (2007) Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand, *Proc. Natl. Acad. Sci. U.S.A. 104*, 5163–5168.
[24] Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage, *Nature 551*, 464–471.
[25] Thuronyi, B. W., Koblan, L. W., Levy, J. M., Yeh, W. H., Zheng, C., Newby, G. A., Wilson, C., Bhaumik, M., Shubina-Oleinik, O., Holt, J. R., and Liu, D. R. (2019) Continuous evolution of base editors with expanded target compatibility and improved activity, *Nat. Biotechnol. 37*, 1070–1079.
[26] Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules, *Nature 472*, 499–503.
[27] Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. (2014) Negative selection and stringency modulation in phage-assisted continuous evolution, *Nat. Chem. Biol. 10*, 216–222.
[28] Boder, E. T., Midelfort, K. S., and Wittrup, K. D. (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity, *Proc. Natl. Acad. Sci. U.S.A. 97*, 10701–10705.
[29] Wang, L., Jackson, W. C., Steinbach, P. A., and Tsien, R. Y. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation, *Proc. Natl. Acad. Sci. U.S.A. 101*, 16745–16749.
[30] Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein, *Nat. Biotechnol. 22*, 1567–1572.
[31] Service, R. F. (2018) Protein evolution earns chemistry Nobel, *Science 362*, 142.
[32] Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins, *Nat. Rev. Genet. 16*, 379–394.

[33] Tanaka, T., Lobato, M. N., and Rabbitts, T. H. (2003) Single domain intracellular antibodies: a minimal fragment for direct in vivo selection of antigen-specific intrabodies, *J. Mol. Biol. 331*, 1109–1120.

[34] Dao, T., Yan, S., Veomett, N., Pankov, D., Zhou, L., Korontsvit, T., Scott, A., Whitten, J., Maslak, P., Casey, E., Tan, T., Liu, H., Zakhaleva, V., Curcio, M., Doubrovina, E., O'Reilly, R. J., Liu, C., and Scheinberg, D. A. (2013) Targeting the intracellular WT1 oncogene product with a therapeutic human antibody, *Sci. Transl. Med. 5*, 176ra133.

[35] Visintin, M., Tse, E., Axelson, H., Rabbitts, T. H., and Cattaneo, A. (1999) Selection of antibodies for intracellular function using a two-hybrid in vivo system, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11723–11728.

[36] Zhou, X., Vink, M., Klaver, B., Berkhout, B., and Das, A. T. (2006) Optimization of the Tet-On system for regulated gene expression through viral evolution, *Gene Ther. 13*, 1382–1390.

[37] Fasan, R. (2012) Tuning P450 enzymes as oxidation catalysts, *ACS Catal. 2*, 647–666.

[38] Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., Wang, P., Malvar, T., and Liu, D. R. (2016) Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance, *Nature 533*, 58–63.

[39] Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., and Liu, D. R. (2013) Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution, *Proc. Natl. Acad. Sci. U.S.A. 110*, 9007–9012.

[40] Packer, M. S., Rees, H. A., and Liu, D. R. (2017) Phage-assisted continuous evolution of proteases with altered substrate specificity, *Nat. Commun. 8*, 956.

[41] Bryson, D. I., Fan, C., Guo, L. T., Miller, C., Soll, D., and Liu, D. R. (2017) Continuous directed evolution of aminoacyl-tRNA synthetases, *Nat. Chem. Biol. 13*, 1253–1260.

[42] Wang, T., Badran, A. H., Huang, T. P., and Liu, D. R. (2018) Continuous directed evolution of proteins with improved soluble expression, *Nat. Chem. Biol. 14*, 972–980.

[43] Roth, T. B., Woolston, B. M., Stephanopoulos, G., and Liu, D. R. (2019) Phage-assisted evolution of *Bacillus methanolicus* methanol dehydrogenase 2, *ACS Synth. Biol. 8*, 796–806.

[44] Salivar, W. O., Tzagoloff, H., and Pratt, D. (1964) Some physical-chemical and biological properties of the rod-shaped coliphage M13, *Virology 24*, 359–371.

[45] Marzio, G., Verhoef, K., Vink, M., and Berkhout, B. (2001) In vitro evolution of a highly replicating, doxycycline-dependent HIV for applications in vaccine studies, *Proc. Natl. Acad. Sci. U.S.A. 98*, 6342–6347.

[46] Verhoef, K., Marzio, G., Hillen, W., Bujard, H., and Berkhout, B. (2001) Strict control of human immunodeficiency virus type 1 replication by a genetic switch: Tet for Tat, *J. Virol. 75*, 979–987.

[47] Badran, A. H., and Liu, D. R. (2015) *In vivo* continuous directed evolution, *Curr. Opin. Chem. Biol. 24*, 1–10.

[48] Tee, K. L., and Wong, T. S. (2013) Polishing the craft of genetic diversity creation in directed evolution, *Biotechnol. Adv. 31*, 1707–1721.
[49] Cupples, C. G., and Miller, J. H. (1989) A set of lacZ mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions, *Proc. Natl. Acad. Sci. U.S.A. 86*, 5345–5349.
[50] Tessman, I., Ishiwa, H., and Kumar, S. (1965) Mutagenic effects of hydroxylamine in vivo, *Science 148*, 507–508.
[51] Greener, A., Callahan, M., and Jerpseth, B. (1997) An efficient random mutagenesis technique using an E. coli mutator strain, *Mol. Biotechnol. 7*, 189–195.
[52] Badran, A. H., and Liu, D. R. (2015) Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra, *Nat. Commun. 6*, 8425.
[53] Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabási, A. L., Oltvai, Z. N., and Osterman, A. L. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655, *J. Bacteriol. 185*, 5673–5684.
[54] Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., and Sabatini, D. M. (2015) Identification and characterization of essential genes in the human genome, *Science 350*, 1096–1101.
[55] Anderson, J. P., Daifuku, R., and Loeb, L. A. (2004) Viral error catastrophe by mutagenic nucleosides, *Annu. Rev. Microbiol. 58*, 183–205.
[56] Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, *Nature 533*, 420–424.
[57] Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., and Kondo, A. (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems, *Science 353*, aaf8729-8721.
[58] Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., Kim, Y. B., Badran, A. H., and Liu, D. R. (2017) 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.
[59] Kim, D., Lim, K., Kim, S. T., Yoon, S. H., Kim, K., Ryu, S. M., and Kim, J. S. (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases, *Nat. Biotechnol. 35*, 475–480.
[60] Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., and Dueber, J. E. (2018) CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window, *Nature 560*, 248–252.
[61] Hess, G. T., Fresard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., Montgomery, S. B., and Bassik, M. C. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells, *Nat. Methods 13*, 1036–1042.
[62] Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells, *Nat. Methods 13*, 1029–1035.

[63] Camps, M., Naukkarinen, J., Johnson, B. P., and Loeb, L. A. (2003) Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I, *Proc. Natl. Acad. Sci. U.S.A. 100*, 9727–9732.
[64] Allen, J. M., Simcha, D. M., Ericson, N. G., Alexander, D. L., Marquette, J. T., Van Biber, B. P., Troll, C. J., Karchin, R., Bielas, J. H., Loeb, L. A., and Camps, M. (2011) Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication, *Nucleic Acids Res. 39*, 7020–7033.
[65] Crook, N., Abatemarco, J., Sun, J., Wagner, J. M., Schmitz, A., and Alper, H. S. (2016) In vivo continuous evolution of genes and pathways in yeast, *Nat. Commun. 7*, 13051.
[66] Ravikumar, A., Arrieta, A., and Liu, C. C. (2014) An orthogonal DNA replication system in yeast, *Nat. Chem. Biol. 10*, 175–177.
[67] Chauthaiwale, V. M., Therwath, A., and Deshpande, V. V. (1992) Bacteriophage lambda as a cloning vector, *Microbiol. Rev. 56*, 577–591.
[68] Hines, J. C., and Ray, D. S. (1980) Construction and characterization of new coliphage M13 cloning vectors, *Gene 11*, 207–218.

Chapter 2: Optimizing Recombineering for the Production of Recombinant Adenoviruses with Multiple and Large DNA Modifications

Portions of the work presented in this chapter have been adapted from the following manuscript and appear in the thesis of Dr. Chet Berman:

Berman, C. M.*; Papa, L. J., III*; Hendel, S. J.*; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D. "An adaptable platform for directed evolution in human cells" *Journal of the American Chemical Society* **2018**, 140, 18093–18103.

*Denotes equal contributions

Other portions of the work presented in this chapter have been adapted from the following manuscript:

Papa, L. J., III; Shoulders, M. D. "Genetic Engineering by DNA Recombineering" *Current Protocols in Chemical Biology* **2019**, 11, e70.

2.1 Author Contributions

L.J.P. designed and performed the recombineering experiments in this chapter. M.D.S. acquired funding and supervised experiments and data analysis.

2.2 Introduction

In this chapter, I present my work optimizing recombineering, a sophisticated homologous recombination-based genetic engineering technique, for the efficient and extensive modification of adenoviral genomes. While we employed recombineering for many purposes beyond adenoviral modification throughout the work presented in this thesis, this technique was particularly pivotal in the work presented in Chapter 3, where we needed to generate genetically modified adenoviral genomes for our human-cell based directed evolution platform. The incorporation of multiple deletions and insertions into the adenoviral genome was greatly complicated by the large size of the ~36 kb adenoviral genome. There are very few unique restriction endonuclease sites in the adenoviral genome, since most restriction endonuclease sites are 6 bp or fewer and are thus likely to occur at least twice by random chance within a ~36 kb genome. Further, the lack of unique restriction sites close to the regions we needed to modify severely complicated efforts to use standard restriction cloning techniques. The difficulty of amplifying large stretches of DNA greater than ~5 kb by PCR[1,2] also complicated the use of DNA assembly methods, since many fragments would be required to assemble the full adenoviral genome and assembly efficiency decreases with the number of fragments.[3-5] However, the most important consideration is that the efficiency of DNA uptake in *Escherichia coli* via heat-shock transformation[6] or electroporation[7] decreases drastically with DNA molecule size, and is extremely low for DNA molecules larger than 10 kb. Thus, in practice, *in vitro* manipulation of the adenoviral genome is not a viable strategy, whether by restriction cloning or by DNA assembly.

In order to overcome these obstacles, we turned to recombineering, which uses small PCR-generated DNA cassettes to make genetic changes to DNA within living *E. coli*.[8] In recombineering, small DNA cassettes are electroporated into the cells with high efficiency while the large Bacterial Artificial Chromosome (BAC) containing the adenoviral genome remains in the cell throughout the editing process, circumventing the inefficiency of transforming large constructs. Recombineering is particularly useful because it relies on the expression of the lambda phage *red* recombination genes, which only require 30–50 bp of homology to carry out homologous recombination.[8] Therefore, recombination cassettes can be generated by simply appending the homology arms as primer overhangs during PCR amplification of the cassette (**Figure 2.1a**). Other homologous recombination-based methods generally require 500–1000 bp homology arms that need to be appended to a targeting cassette through time-consuming restriction cloning or difficult overlap extension PCR. [9-12]

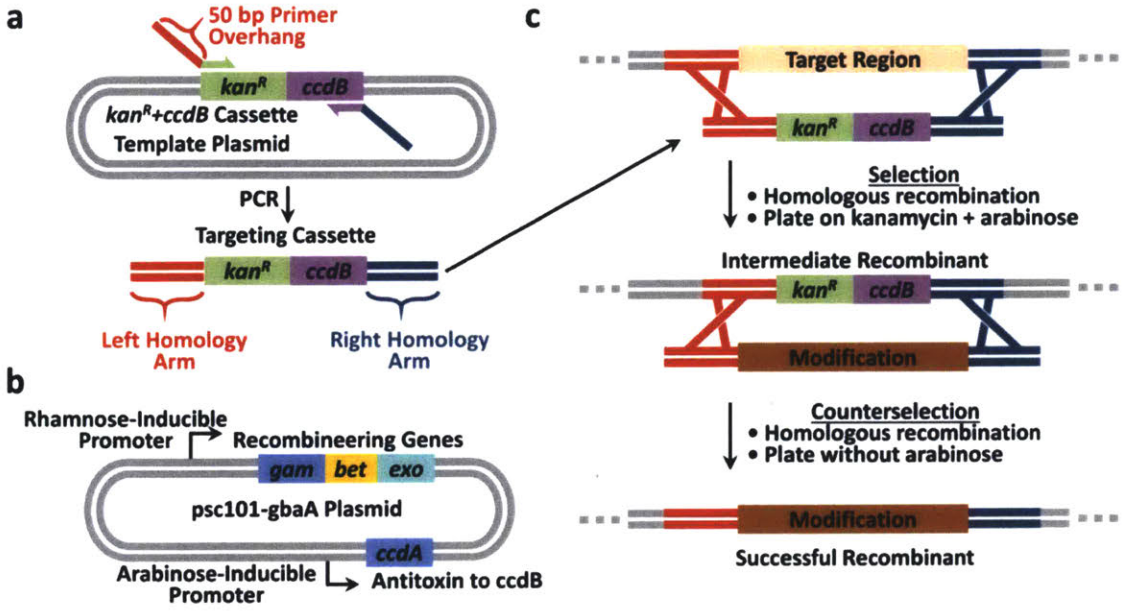

Figure 2.1 Overview of recombineering using lambda *red* genes.
(**a**) Targeting cassettes are generated by amplifying the desired modification or selection cassette using primers that have short homology arm overhangs. (**b**) Recombineering is carried out using a plasmid that inducibly expresses the lambda *red* genes when L-rhamnose is present, and that inducibly expresses the ccdA antitoxin when arabinose is present. (**c**) In the first step of recombineering, a *kan*^{R}+*ccdB* cassette is inserted into the region to be modified and selected for on kanamycin plates. In the second step, the *kan*^{R}+*ccdB* genes are replaced by the final modification, and unmodified cells are counterselected against by removing arabinose, and thus the ccdA antitoxin, and rendering the ccdB toxin lethal.

Each modification made using recombineering is generally a two-step process, with the lambda *red* recombination genes required for each recombineering step being expressed from the psc101-gbaA recombineering plasmid via L-rhamnose induction (**Figure 2.1b**).[13] In the first step, a DNA cassette containing a selectable and a counterselectable marker is targeted to the region being modified or replaced (**Figure 2.1c**). Under standard recombineering conditions with double stranded DNA cassettes, only 1 in 10,000–100,000 cells yields successful recombinants that incorporate the targeting cassette at the desired location.[14] This frequency is far too low to reasonably screen colonies one-by-one for recombinants, which is why one must use a selection marker—we chose to use a kanamycin resistance gene to make our modifications—to select for rare successful recombinants. While the ampicillin resistance gene (β-lactamase) is a more common selectable marker for recombineering,[13,15,16] we settled on using the kanamycin resistance gene instead. The β-lactamase enzyme that confers ampicillin resistance is secreted and destroys extracellular antibiotic, allowing some nearby susceptible cells to survive and grow,[17] which could cause unwanted non-recombinant contamination in the second recombineering step.

The kanamycin-resistant recombinants generated after the first selection step are only intermediates. The kanamycin resistance gene is removed along with the counterselection in a second recombination step using a targeting DNA cassette that

completes the final desired insertion, deletion or replacement (**Figure 2.1c**). Rare recombinants of the second step are selected for by rendering the counterselectable marker toxic—we chose to use the ccdB toxin—such that only the rare few cells that have replaced the kanamycin resistance gene and the counterselectable marker with the final modification will survive. The *ccdB* gene encodes for a toxic protein inhibitor of gyrase in *E. coli,* leading to cell death unless the antitoxin protein encoded by the *ccdA* gene binds and inhibits the ccdB toxin.[13] Throughout the recombineering protocol, the ccdA antitoxin is expressed by arabinose induction from the psc101-gbaA recombineering plasmid to render *ccdB* non-lethal until the counterselection step (**Figure 2.1b**). After the second recombineering step, arabinose is removed so that ccdA antitoxin is no longer expressed and any cells still containing *ccdB* die, thus achieving counterselection.

Out of the many toxic counterselection genes available, we chose a protocol that utilizes the *ccdB* counterselection gene[13] because it reliably kills *ccdB*-containing cells and counterselection only takes a day. With the popular *galK* counterselection protocol,[18] minimal media plates need to be incubated for a few days to observe colonies. Similarly, while there are quicker one-step recombineering protocols that introduce a selectable marker, such as a kanamycin resistance gene, as a permanent part of the final modification,[19] we decided against this protocol for two main reasons. First, we did not want the kanamycin resistance gene to be present in the final adenoviral genome because we wanted to reserve as much of the ~38 kb capacity of the adenoviral capsid[20] for the BOI as possible. Second, we needed to make multiple modifications and thus needed to reuse the kanamycin selection marker several times.

Our initial attempts to modify the adenoviral genome using recombineering were hampered by an unexpectedly high rate of target DNA multimerization. We found that this phenomenon is rarely mentioned in the recombineering literature, but greatly impacted our efforts to carry out the required multi-step series of modifications. We also found that it was important to purify large >2 kb PCR cassettes to prevent unwanted recombination with short, truncated PCR side products. After modifying our recombineering workflow to account for a high rate of target multimerization and for PCR side products, we were able to successfully generate the highly-modified adenovirus needed to realize the human cell-based continuous directed evolution platform presented in Chapter 3.

2.3 Results and Discussion

2.3.1 Recombineering-induced BAC multimerization

While recombineering has been used routinely for making deletions or insertions in large viral genomes such as cytomegalovirus[15] and adenovirus,[16,21-23] our initial attempts to edit the adenoviral genome were unsuccessful. Often, we were perplexed to find that all of the colonies that survived counterselection contained unmodified adenoviral genomes. We also found that on the rare occasion we could successfully make one modification, we would be unsuccessful when trying to make a second modification on the same adenoviral genome. This appearance of a high number of false positives after selection and/or counterselection has been noted in the recombineering literature as usually being caused by unintentional transformation of the selection-counterselection template plasmid.[24,25] After generating the selection-counterselection targeting cassette (**Figure 2.1a**), it is vital to remove or destroy any of the remaining template plasmid after the PCR by gel extraction and/or DpnI digestion.[24,25] Even a miniscule amount of intact template plasmid contamination in the targeting cassette preparation will result in a much larger number of colonies upon selection that have taken up the template plasmid than the small number of colonies that have successfully recombined the targeting cassette into the target. Colonies that have survived selection by taking up the template plasmid rather than undergoing successful recombination, can subsequently survive the counterselection step by simply losing the template plasmid. As an aside, it is important to note that spontaneous plasmid loss happens at a much higher rate than successful recombination, and unless plasmid loss is selected against, it is the most frequent outcome of counterselection. We selected against plasmid loss with the antibiotic chloramphenicol, since the BAC containing the adenoviral genome also carried a chloramphenicol resistance gene.

One common strategy for preventing template plasmid contamination in recombineering is to use template plasmids with R6K origins, which are replication incompetent in most *E. coli* strains.[24] The R6K origin can only replicate in cells expressing the pi protein,[26] such as Pir2 *E. coli* cells, and thus cannot confer antibiotic resistance to the DH10B cells we used for recombineering. However, after we created a new *kanR+ccdB* template plasmid utilizing the R6K origin, R6K-kan-ccdB (Genbank Accession No. MH325106), we surprisingly still observed a very high rate of false positive colonies after counterselection.

We surmised that if multiple copies of the target BACs were recombining to form multimers, such as dimers or trimers, this could also result in a high rate of false positives. For example, if a dimer were to undergo recombination to form a heterodimer of an unmodified target BAC and a *kan^R+ccdB* modified target BAC in order to survive selection, the heterodimer could then undergo intramolecular recombination to yield a monomer of the unmodified target BAC again to survive counterselection (**Figure 2.2**). It has been noted that recombineering can cause multi-copy plasmids with pBR322/colE1-derived origins to recombine into multimers,[27-29] and indeed the formation of multimeric circular lambda phage genomes is one of the proposed natural functions of the lambda *red* genes.[8,30] Interestingly, previous work found that plasmids with low-copy psc101 origins did not appear to form multimers upon recombineering.[29] However, most recombineering is not done with multi-copy plasmid origins such as pBR322/colE1 or

psc101, but rather with BACs, which utilize the single-copy origin of the F-plasmid.[31] BACs actively partition during division,[32,33] thus preventing mixtures of unmodified monomeric parent BACS and modified monomeric BACs from forming within one cell after recombineering. The formation of BAC multimers has been alluded to once,[34] but any warning about this phenomenon is absent in the vast majority of BAC recombineering protocols.

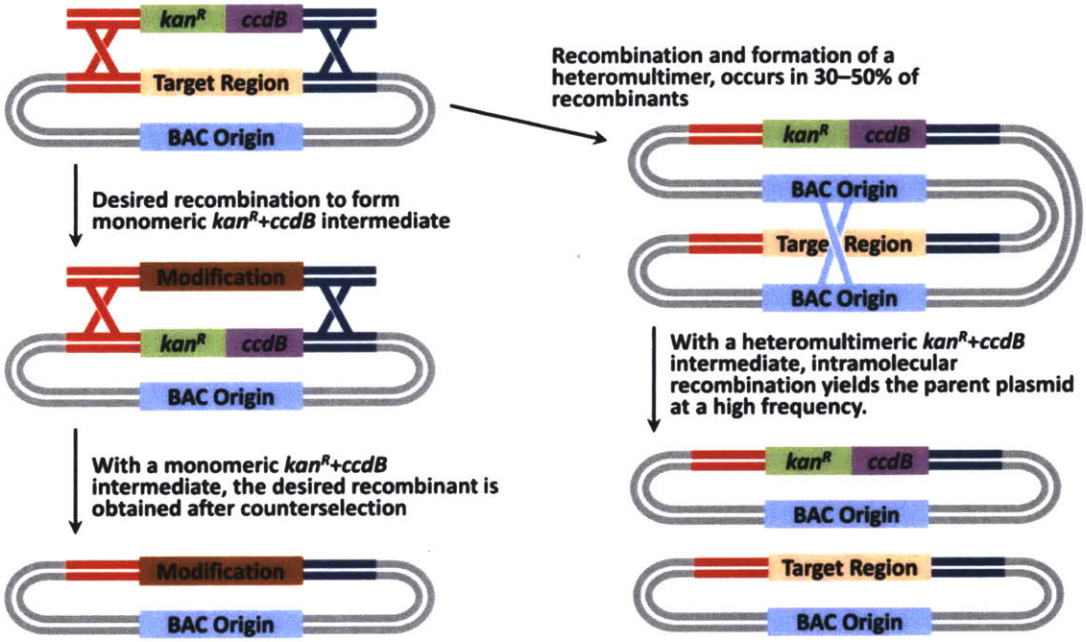

Figure 2.2 Heteromultimeric *kan*^{R}*+ccdB* intermediates generate false positives. Recombineering with monomeric *kan*^{R}*+ccdB* intermediates (left) generally yields successful recombinants because cells must replace the *ccdB* gene to survive counterselection. Heteromultimers of the *kan*^{R}*+ccdB* intermediate and unmodified parent BAC (right) undergo intramolecular recombination at a high rate to eliminate the toxic *ccdB* gene, yielding unmodified parent BAC at a frequency much higher than successful recombineering.

To investigate the extent to which our adenoviral genome target BACs might be multimerizing, we conducted a simple experiment in which we attempted to insert a *kan*^{R}*+ccdB* cassette into a *lacZα*-containing adenoviral genome BAC (**Figure 2.3a**). If recombineering successfully interrupts the *lacZα* gene via *kan*^{R}*+ccdB* insertion without forming a heteromultimer, the resulting colonies will not turn blue on plates containing kanamycin and X-gal due to the loss of *lacZα*, which metabolizes the colorless X-gal substrate into a dark blue product. However, if a heteromultimer is formed, the resulting colony will contain both a kanamycin resistance gene and an uninterrupted *lacZα* gene, resulting in dark blue kanamycin resistant colonies. When we performed this recombineering, we were shocked to find that ~40% of the colonies were blue, suggesting a very high rate of heteromultimer formation (**Figure 2.3b**). Furthermore, we found that when we attempted to replace the *kan*^{R}*+ccdB* cassette of one of the blue, presumably heteromultimeric colonies, with *rtTA*, which does not metabolize X-gal, we obtained a very

high number of almost only blue colonies, with 3 out of every 1000 cells surviving counterselection (**Figure 2.3c**), a rate much higher than the number of recombinants normally obtained for recombineering,[14] 1 in 10,000 to 1 in 100,000. The very high rate of blue colonies is consistent with our hypothesis that the heteromultimers escape counterselection by resolving back into the parent monomer without undergoing the desired modification (**Figure 2.2**).

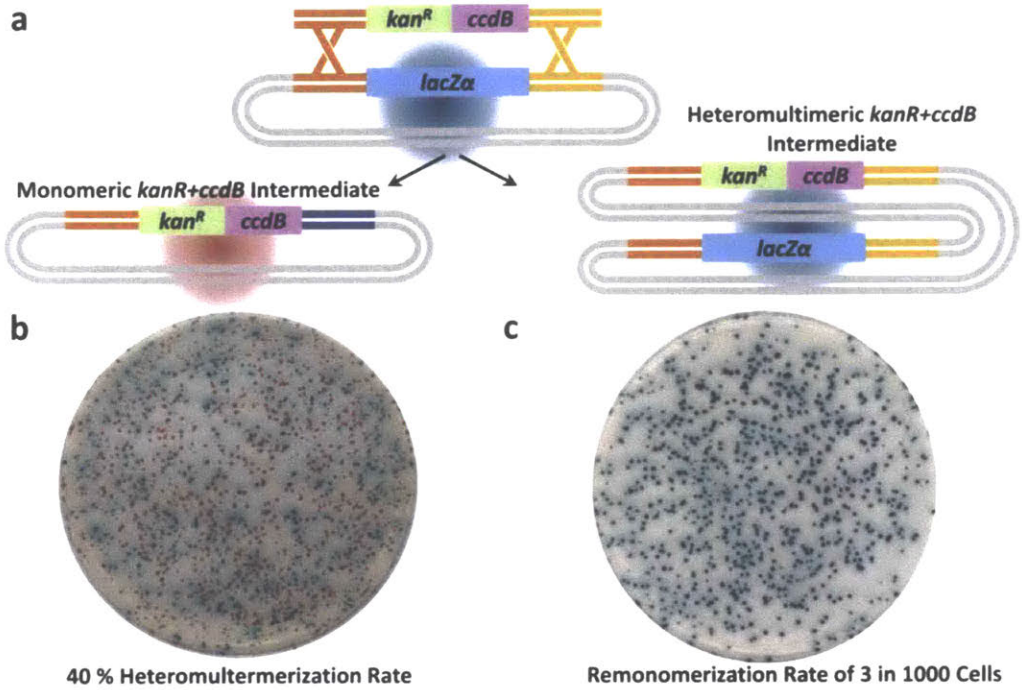

Figure 2.3 Adenoviral genome BACS heteromultimerize at a high rate.
(**a**) In an assay to assess the rate of heteromultimerization, the *lacZα* gene is replaced by *kan*^{R}*+ccdB*. If a monomeric *kan*^{R}*+ccdB* intermediate forms, the colony will not be able to metabolize X-gal into a blue substrate, and will simply turn red from the tetrazolium chloride dye on the plate. If a heteromultimeric *kan*^{R}*+ccdB* intermediate forms, the colony will both survive kanamycin selection and metabolize X-gal into a blue substrate. (**b**) When attempting to replace the *lacZα* with *kan*^{R}*+ccdB*, ~40% of the resulting recombinants are still blue, and likely heteromultimeric. (**c**) Attempting to replace *kan*^{R}*+ccdB* in a putative heteromultimer results in a high number of blue colonies. The plate in panel **c** is diluted 10-fold relative to the number of cells plated in panel **b**.

Taking into account our new awareness about the surprisingly high multimerization rate, we were finally able to delete a variety of adenovirus genes by picking three or four intermediates, we ensured that at least one of the colonies picked was a monomer rather than a heteromultimer. Even without sequencing the results of counterselection, it was often very clear which counterselections were the results of heteromultimers remonomerization to parent BACs versus successful recombineering with a *kan*^{R}*+ccdB* monomer because the former produced significantly more colonies than the latter. We were able to detect heteromultimers with different techniques, such as analyzing restriction digest patterns or by performing colony PCR on the region being modified,[35] however we found that it was often easier and faster to blindly proceed with 3 random

kan^R+*ccdB* intermediates than to take the time to screen for one *kan*^{R}+*ccdB* monomer. Once it was clear which of the three *kan*^{R}+*ccdB* intermediates yielded successful recombinants, which was often obvious from the number of colonies resulting from counterselection, we would make a glycerol stock of that intermediate as a "verified" monomer so that we could make different modifications in the same location in the future.

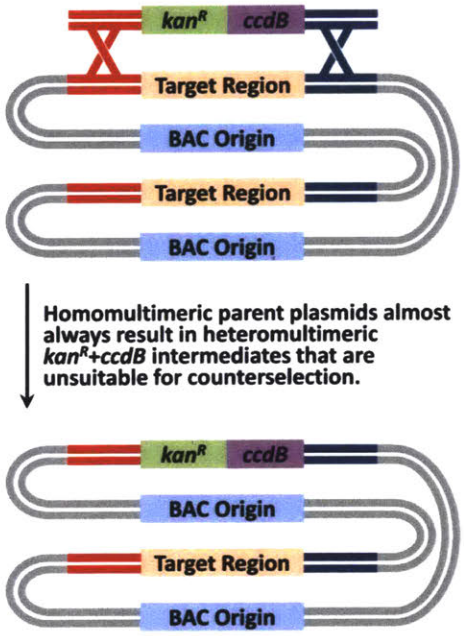

Figure 2.4 Homomultimeric parent BACs yield heteromultimeric *kan*^{R}+*ccdB* intermediates.
BACs that are homomultimeric will almost always form heteromultimeric *kan*^{R}+*ccdB* intermediates, which are unsuitable for counterselection owing to the high rate of intramolecular recombination to yield unmodified parent BAC.

Occasionally, we found that after making an initial deletion, we were unable to make any more deletions no matter how many *kan*^{R}+*ccdB* intermediates we tested. In these cases, instead of ~40% of *kan*^{R}+*ccdB* intermediates being heteromultimeric, 100% of the intermediates were heteromultimeric. The high rate of heteromultimer formation is likely a result of the parent vector being a homomultimer, since most recombineering events will only happen in one of the multiple adenoviral genomes in a homomultimer (**Figure 2.4**). Hence, if a homomultimer forms after a successful deletion, it will be an unsuitable starting vector for making another deletion. One potential solution is to digest purified homomultimers into monomers with a restriction enzyme that cuts each multimer once, and to then relegate the monomers and transform them back into cells.[34] However, because our adenoviral genome BACs are so large (~40 kb), transformation efficiency was extremely low and we were rarely able to remonomerize multimers of the adenoviral genome in general. Additionally, it was very difficult to assess whether remonomerization

was successful, since homomultimers will give the same gel patterns as monomers after a digestion or colony PCR test,[35] and since special gel electrophoresis equipment and techniques are required to reliably resolve plasmids larger than ~20 kb.[36]

Since our new recombineering workflow involved picking three or four *kan*^{R}+*ccdB* intermediates, each counterselection step informed us about the heteromultimerization rate of the previous selection step. Normally at least one of the three or four plates would be sparsely populated with mostly successful recombinants, while the other plates would be densely populated, indicative of heteromultimer remonomerization (**Figure 2.5a**). If all three or four plates after counterselection were densely populated, it was suggestive that the parent BAC preceding the selection step was homomultimeric (**Figure 2.5b**). We found that the most reliable and simplest way to detect potential homomultimers was by observing this 100% heteromultimerization as indicated by all counterselections resulting in densely populated plates. Because of the great effort required in detecting and remonomerizing homomultimers, we simply discarded any suspected homomultimers if all three or four *kan*^{R}+*ccdB* intermediates generated from the suspected homomultimer yielded a very large number of colonies after counterselection, especially after a second attempt with 4 more *kan*^{R}+*ccdB* colonies (**Figure 2.5b**). We then would redo the previous recombineering step or go back to the counterselection plate that the suspected homomultimer came from, and pick a new colony with the hopes that it is a monomer and thus a suitable starting plasmid for the next modification. Homomultimers seemed to be fairly rare and form with less frequency than heteromultimeric *kan*^{R}+*ccdB* intermediates, so we would still only proceed with one colony at a time after counterselection, unlike the three or four colonies we would proceed with after selection.

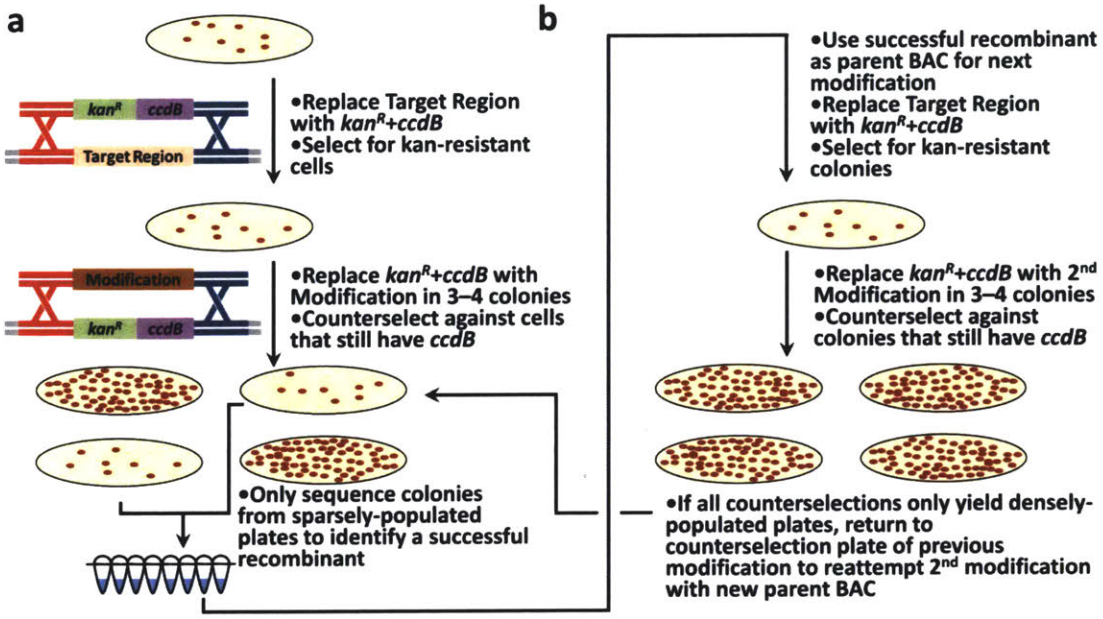

Figure 2.5 Efficient recombineering workflow to mitigate multimer formation. (**a**) An example of a successful recombineering workflow resulting from a monomeric parent BAC. At least some of the *kanR+ccdB* intermediates used to make the final modification should be monomeric, and thus yield a small number of colonies. (**b**) An example of an unsuccessful recombineering workflow resulting from a homomultimeric parent BAC. All *kanR+ccdB* intermediates are heteromultimeric, and thus result only in densely populated plates after counterselection due to the high rate of heteromultimerization.

It is not clear why the high rate of BAC multimerization is not explicitly mentioned in the many recombineering protocols that have been published. Perhaps some specific property of the adenoviral genome BAC makes it more prone to multimerization than most BACs or the particular recombineering protocol or bacterial strain we utilized induces multimer formation more readily than other protocols or strains. A much more likely reason is that, unlike our adenoviral BACs that require multiple modifications made in series, many researchers only make a single modification to a BAC or a set of single modifications in parallel, meaning that the formation of homomultimers is no longer problematic since there are no subsequent modifications to be made via recombineering. It is also worth noting that many researchers use recombineering to directly make modifications to the *E. coli* chromosome itself, which has its own dedicated system for resolving genome multimers back into monomers before the cell divides,[37] thus ensuring that all *kan*^{R}*+ccdB* intermediates and recombinant products are monomeric. Regardless, once we recognized multimerization as the principle cause for our recombineering challenges, we successfully deleted various adenovirus genes in series, such as AdPol, AdProt, Fiber, DBP, pTP, and PVI.

2.3.2 Optimizing the insertion of large DNA cassettes

Despite our modified workflow allowing for the reliable deletion of multiple genes in series (**Figure 2.5**), we had limited success inserting various BOI genes and fluorescent proteins into the adenoviral genome. While the region of interest would frequently be modified with at least a portion of the desired insertion, the insertions almost always contained random internal deletions (**Figure 2.6**).

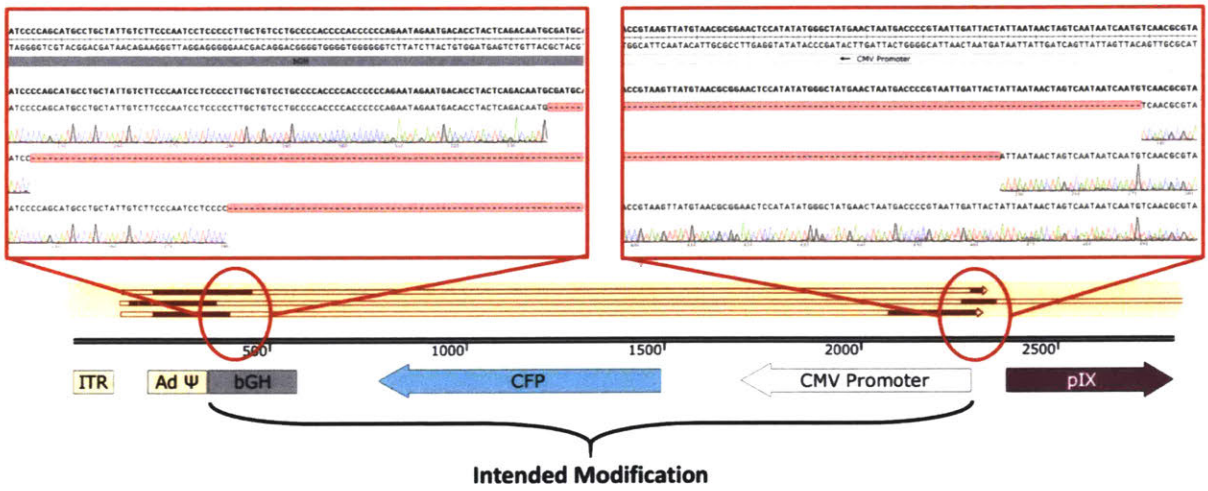

Figure 2.6 Attempted insertions containing unwanted internal deletions.
Representative Sanger sequencing traces (*top*) of recombinants from an attempted insertion showing the junctions of unwanted internal deletions, which are apparent in the read alignments (*bottom*). Images generated using SnapGene® software (GSL Biotech).

Interestingly, we found that purifying the targeting cassette by gel extraction would largely eliminate these unwanted internal deletions (**Figure 2.7**). We suspect that gel extraction removes small truncated PCR side products that recombine into the target at a much higher efficiency than the desired full-length targeting cassette. Indeed, the efficiency of recombineering is known to depend on targeting cassette size, with smaller cassettes recombining at a higher rate than larger cassettes, and cassettes larger than 2.5 kb recombineering at a very low efficiency.[38]

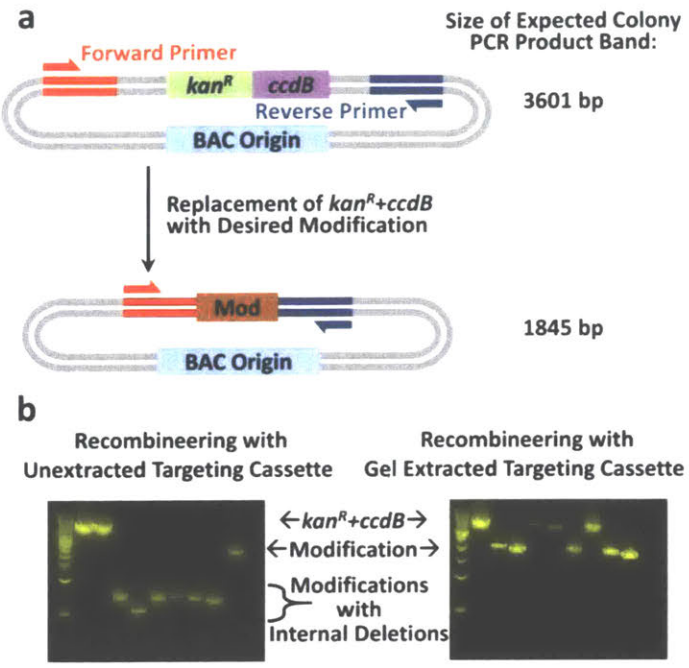

Figure 2.7 Gel extracted targeting cassettes increases the rate of successful insertion.
(**a**) Colony PCR is performed using primers that flank the region being modified. Because the modification significantly changes the distance between the primers, the difference between the sizes of amplicons generated from the *kan*^{R}*+ccdB* intermediate, the desired modification, and modifications containing internal deletions is apparent on an agarose gel. (**b**) Agarose gels showing the colony PCR results from 8 random colonies after recombination with an unextracted (*left*) or gel extracted (*right*) targeting cassette. Each gel shows from left to right: 1 kb ladder (New England Biolabs B7025), colony PCR amplicon of a *kan*^{R}*+ccdB* control, and colony PCR amplicons of 8 random colonies after counterselection.

Previous recombineering protocols generally recommend gel extraction as a means of removing template plasmids from targeting cassette preparations,[18,25,39] however we find that even with non-replicating template plasmids it is still important to gel extract the targeting cassette if it is larger than 2 kb, since it could easily be outcompeted by smaller PCR side products. It is important to note that UV transilluminators can cause DNA damage and unwanted mutations, which is why some protocols recommend using non-UV DNA visualization techniques or avoiding the need to gel extract altogether with non-replicating templates.[39,40] We opted to use a blue-LED transilluminator and GelGreen® dye rather than a UV transilluminator and ethidium bromide dye. While gel extracting large cassettes eliminated the high background of recombinants that incorporated short PCR side products, it did not improve the intrinsically low recombination rate of the large cassettes. Very low efficiency recombinations with large

cassettes often did not produce enough successful recombinants after counterselection relative to the low background of cells that mutated and inactivated the *ccdB* gene. For particularly low efficiency recombinations, increasing the number of cells being electroporated from a 1 mL to a 25 mL scale resulted in a high enough rate of successful recombinants to reasonably screen for among the background of *ccdB* mutants.

2.4 Conclusion

We found that that the reported lambda *red* recombineering technique had many pitfalls owing to successful recombination being an intrinsically rare event compared to the multiple competing processes that occurred more frequently. Once we developed and implemented reliable measures to mitigate or eliminate each of these competing processes, particularly multimerization of the target BAC and unwanted recombination with truncated PCR side products, we were able to efficiently generate a large panel of adenoviruses, each with multiple deletions and insertions. The lessons learned in our initial efforts to edit the adenoviral genome proved invaluable not only for the work presented in Chapter 3, but also for the work presented in Chapters 4 and 5, in which we extensively engineered the *E. coli* genome and the lambda phage genome, respectively.

2.5 Materials and Methods

General methods: All PCR reactions for generating recombineering targeting cassettes were performed using Q5 High Fidelity DNA Polymerase (New England BioLabs). Primers were obtained from Life Technologies and Sigma-Aldrich (**Table 2.1**).

Table 2.1 Primers used for recombineering.

Name	Sequence
E1.CMV.Promoter.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccacccttaagccacgcccacaga tatacgcgttgacattg-3′
E1.bGH.polyA.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaatagaagccataga gcccac-3′
LacZ.kanccdB.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccacccttaagccacgcccaccct catcagtgccaacatagtaag-3′
LacZ.kanccdB.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaataccgctcattagg cgggc-3′
ColonyPCR.Forward	5′-tgacgtagtagtgtggcggaagtg-3′
ColonyPCR.Reverse	5′-cggcggctgctgcaaaac-3′

General recombineering protocol: Adenoviral constructs were engineered using *ccdB* recombineering, as previously described[13] and further modified by us,[35] in DH10B *Escherichia coli* carrying the adenovirus type-5 genome in a chloramphenicol-resistant bacterial artificial chromosome (AdBAC). Cells carrying an AdBAC were transformed with the temperature-sensitive psc101-gbaA recombineering plasmid,[13] plated on LB (Difco) agar (Alfa Aesar) with 10 μg/mL tetracycline (CalBioChem) and 10 μg/mL chloramphenicol (Alfa Aesar), and incubated for 24 h at 30 °C. Colonies were selected and grown in LB containing 10 μg/mL tetracycline and 10 μg/mL chloramphenicol overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 μg/mL tetracycline and 10 μg/mL chloramphenicol 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 L-arabinose (Chem-Impex) and L-rhamnose (Sigma Aldrich) 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 *kanR+ccdB* targeting cassette (1.8 kV, 5.8 msec, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in super optimal broth with catabolite repression (SOC; Teknova) with 2 mg/mL L-arabinose at 30 °C for 2 h, then plated on LB agar plates with 50 μg/mL kanamycin (Alfa Aesar) and 2 mg/mL L-arabinose and incubated for 24 h at 30 °C. Colonies that grew under these conditions had incorporated the *kanR+ccdB* targeting cassette and were picked in triplicate and grown in LB with 50 μg/mL kanamycin and 2 mg/mL L-arabinose at 30 °C for 18–21 h. Note that the colonies were picked in triplicate because multimers of the AdBAC formed at a high rate (~30–50% of colonies) during the first recombineering step. Such multimers cannot be successfully recombineered in the next step. Picking three colonies and recombineering them separately in parallel increases the chances of picking a monomer that can be successfully recombineered. The cultures were then diluted 25-fold in LB with 50 μg/mL kanamycin and 2 mg/mL L-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 L-

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 *kanR+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 L-arabinose at 30 °C for 2 h, and then were washed once with LB to remove the L-arabinose and prevent continued production of the ccdA antitoxin. The cultures were then plated on LB agar plates at various dilutions with 10 µg/mL tetracycline and 10 µg/mL chloramphenicol 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, in principle, have the desired final targeting cassette integrated, but were always screened by PCR or sequencing to confirm cassette integration as some colonies may simply inactivate the ccdB toxin.

Rate determination for BAC heteromultimerization and remonomerization: DH10B cells containing the Ad.LacZ BAC (see Appendix B.1 for sequence) and psc101-gbaA grown from a colony in LB containing 10 µg/mL tetracycline and 10 µg/mL chloramphenicol overnight at 30 °C (18 h). The overnight culture was diluted 25-fold in LB with 10 µg/mL tetracycline and 10 µg/mL chloramphenicol in a 1.5 mL Eppendorf tube with a hole in the lid. The culture was grown at 30 °C for 2 h in a shaking heat block at 1000 r.p.m. The ccdA antitoxin and recombineering machinery were then induced by adding L-arabinose and L-rhamnose to a final concentration of 2 mg/mL each and raising the temperature to 37 °C for 40 min. The culture was then placed on ice, washed twice with ice-cold, sterile ddH_2O, resuspended in ~25 µL of ice-cold, sterile ddH_2O, and electroporated (1.8 kV, 5.9 msec, 0.1 cm cuvette, BioRad Micropulser) with 200 ng of a *kanR+ccdB* cassette targeting *lacZα*. The cassette was generated with primers LacZ.kanccdB.Forward and LacZ.kanccdB.Reverse (**Table 2.1**) and template plasmid R6K-kan-ccdB (Genbank Accession No. MH325106), and was purified using the E.Z.N.A. cycle pure kit (Omega Bio-tek) according to the manufacturer's instructions, except that the cassette was eluted into a 0.2× dilution of elution buffer. The cells were then recovered in 1 mL super optimal broth with catabolite repression (SOC; Teknova) with 2 mg/mL L-arabinose at 30 °C for 2 h in the shaking heat block, then 50 µL of the recovery was plate on an LB agar plate with 50 µg/mL kanamycin (Alfa Aesar), 2 mg/mL L-arabinose, 0.2 mg/mL 5'-bromo-4-chloro-3-indolyl-B-galactopyranoside (X-gal; IBI Scientific), 0.24 mg/mL isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma-Aldrich) and 50 µg/mL tetrazolium chloride (TTC; Sigma-Aldrich). The tetrazolium chloride is a metabolic dye that allows for better contrast during plate imaging.[41] The plates were incubated for 48 h at 30 °C then imaged by inverting the plates onto transparencies and scanning on a document scanner at a resolution of 600 dots per inch. The colonies were then counted using the software OpenCFU (3.9.0),[42] with the minimum colony radius set to 3, the maximum colony radius set to 50, and the regular threshold set to 4. Detailed colony count reports were exported as .csv files. Blue and red colonies were filtered in excel according to hue (hue from 340 to 40 assigned as red, other assigned as blue) in order to quantify red and blue colonies separately. The number of blue colonies (776) was

divided by the number of total colonies (1818) to calculate a heteromultimerization rate of 43%.

One of the heteromultimeric blue colonies was picked and grown overnight in LB with 50 µg/mL kanamycin and 2 mg/mL L-arabinose at 30 °C for 18 h. The overnight culture was then diluted 25-fold in LB with 50 µg/mL kanamycin and 2 mg/mL L-arabinose and grown at 30 °C for 2 h. The recombineering machinery was then induced by adding L-rhamnose to a final concentration of 2 mg/mL and then growing the cultures at 37 °C for 40 min. The culture was then placed on ice, washed twice with ice-cold, sterile ddH_2O, resuspended in ~25 µL of ice-cold, sterile ddH_2O, and electroporated (1.8 kV, 5.9 msec, 0.1 cm cuvette, BioRad Micropulser) with 163 ng of a cassette that replaces *kan*^{R}+*ccdB* with rtTA. The cassette was generated with primers E1.CMV.Promoter.Forward and E1.bGH.polyA.Reverse (**Table 2.1**) and template plasmid pCDNA3.1-rtTA (see Appendix B.1 for sequence), and was purified using the E.Z.N.A. cycle pure kit (Omega Bio-tek) according to the manufacturer's instructions, except that the cassette was eluted into a 0.2× dilution of elution buffer. The cells were then recovered in SOC with 2 mg/mL L-arabinose at 30 °C for 2 h, and then pelleted at 10,000 × g and resuspended in 1 mL LB to remove the L-arabinose. 50 µL of a 10^{-1} dilution of the resuspension was plated on an LB agar plate with 10 µg/mL tetracycline, 10 µg/mL chloramphenicol, 0.2 mg/mL X-gal, 0.24 mg/mL IPTG and 50 µg/mL tetrazolium chloride (TTC; Sigma-Aldrich). 50 µL of a 10^{-4} dilution of the resuspension was plated on an LB agar plate with 50 µg/mL kanamycin, 2 mg/mL L-arabinose, and 50 µg/mL tetrazolium chloride. The plates were incubated for 36 h at 30 °C then imaged and counted as above. The chloramphenicol/X-gal/IPTC/TTC plate had 925 blue colonies and 14 red colonies and the kanamycin/arabinose/TTC plate had 296 colonies, indicating a remonomerization rate of 3.1×10^{-3}.

Comparison of recombineering experiments by colony PCR: Recombineering was performed as mention above to replace *kan*^{R}+*ccdB* with transthyretin using a targeting cassette generated using the primers E1.CMV.Promoter.Forward and E1.bGH.polyA.Reverse (**Table 2.1**) and the template plasmid pCDNA3.1-TTR (see Appendix B.1 for sequence). Recombineering was performed in parallel with a targeting cassette that was simply purified using an E.Z.N.A. Cycle Pure kit (Omega Bio-tek) or a targeting cassette that was gel purified using an E.Z.N.A. Gel Extraction kit (Omega Bio-tek). The resulting colonies were picked with a sterile 10 µL pipette tip and each swirled in a PCR tube containing 25 µL of 1× OneTaq® Quick-Load® 2X Master Mix (New England Biolabs) and the primers ColonyPCR.Forward and ColonyPCR.Reverse (**Table 2.1**). The PCRs were performed using the following thermocycler PCR protocol and visualized by gel electrophoresis:

1. 94 °C 5 min
2. 94 °C 15 sec
3. 68 °C 15 sec
4. 68 °C 4 min
5. Return to step 2 for 30 cycles
6. 68 °C 5 min
7. 4 °C hold

2.6 Acknowledgements

This work was supported by an NIH Director's New Innovator Award (1DP2GM119162), MISTI Global Seeds Fund, and MIT (all to M.D.S.) and by the European Union through the 6th Framework Program GIANT (contract no. 512087; to R.C.H.). L.J.P., C.L.M., and S.J.H. were supported by National Science Foundation Graduate Research Fellowships under Grant No. 1122374. C.M.B. was supported by an NIGMS/NIH Inter-Departmental Biotechnology Training Program (T32-GM008334). This work was also supported in part by the NIH/NIEHS under award P30-ES002109 and by Cancer Center Support (core) Grant P30-CA14051 from the NIH/NCI.

2.7 References

[1] Cheng, S., Fockler, C., Barnes, W. M., and Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA, *Proc. Natl. Acad. Sci. U.S.A. 91*, 5695–5699.

[2] Chua, E. W., Miller, A. L., and Kennedy, M. A. (2015) Choice of PCR microtube can impact on the success of long-range PCRs, *Anal. Biochem. 477*, 115–117.

[3] Birla, B. S., and Chou, H. H. (2015) Rational design of high-number dsDNA fragments based on thermodynamics for the construction of full-length genes in a single reaction, *PLoS One 10*, e0145682.

[4] (2014) *DNA Cloning and Assembly Methods*, Vol. 1116, Springer.

[5] Richter, D., Bayer, K., Toesko, T., and Schuster, S. (2019) ZeBRalpha a universal, multi-fragment DNA-assembly-system with minimal hands-on time requirement, *Sci. Rep. 9*, e0145682.

[6] Inoue, H., Nojima, H., and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids, *Gene 96*, 23–28.

[7] Siguret, V., Ribba, A. S., Chérel, G., Meyer, D., and Piétu, G. (1994) Effect of plasmid size on transformation efficiency by electroporation of *Escherichia coli* DH5α, *BioTechniques 16*, 422–426.

[8] Murphy, K. C. (2016) λ recombination and recombineering, *EcoSal Plus 7*.

[9] Winans, S. C., Elledge, S. J., Krueger, J. H., and Walker, G. C. (1985) Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*, *J. Bacteriol. 161*, 1219–1221.

[10] Kong, Y., Yang, T., and Geller, A. I. (1999) An efficient in vivo recombination cloning procedure for modifying and combining HSV-1 cosmids, *J. Virol. Methods 80*, 129–136.

[11] Uil, T. G., Vellinga, J., de Vrij, J., van den Hengel, S. K., Rabelink, M. J., Cramer, S. J., Eekels, J. J., Ariyurek, Y., van Galen, M., and Hoeben, R. C. (2011) Directed adenovirus evolution using engineered mutator viral polymerases, *Nucleic Acids Res. 39*, e30.

[12] Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P., and Kushner, S. R. (1989) New method for generating deletions and gene replacements in *Escherichia coli*, *J. Bacteriol. 171*, 4617–4622.

[13] Wang, H., Bian, X., Xia, L., Ding, X., Muller, R., Zhang, Y., Fu, J., and Stewart, A. F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection, *Nucleic Acids Res. 42*, e37.

[14] Datta, S., Costantino, N., Zhou, X., and Court, D. L. (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages, *Proc. Natl. Acad. Sci. U.S.A. 105*, 1626–1631.

[15] Murrell, I., Wilkie, G. S., Davison, A. J., Statkute, E., Fielding, C. A., Tomasec, P., Wilkinson, G. W., and Stanton, R. J. (2016) Genetic stability of bacterial artificial chromosome-derived human cytomegalovirus during culture *in vitro*, *J. Virol. 90*, 3929–3943.

[16] Muck-Hausl, M., Solanki, M., Zhang, W., Ruzsics, Z., and Ehrhardt, A. (2015) Ad 2.0: a novel recombineering platform for high-throughput generation of tailored adenoviruses, *Nucleic Acids Res. 43*, e50.

[17] Medaney, F., Dimitriu, T., Ellis, R. J., and Raymond, B. (2016) Live to cheat another day: bacterial dormancy facilitates the social exploitation of beta-lactamases, *ISME J. 10*, 778–787.

[18] Warming, S., Costantino, N., Court, D. L., Jenkins, N. A., and Copeland, N. G. (2005) Simple and highly efficient BAC recombineering using *galK* selection, *Nucleic Acids Res. 33*, e36.

[19] Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proc. Natl. Acad. Sci. U.S.A. 97*, 6640–6645.

[20] Bett, A. J., Prevec, L., and Graham, F. L. (1993) Packaging capacity and stability of human adenovirus type 5 vectors, *J. Virol. 67*, 5911–5921.

[21] Zhang, W., and Ehrhardt, A. (2017) Getting genetic access to natural adenovirus genomes to explore vector diversity, *Virus Genes 53*, 675–683.

[22] Zhang, W., Fu, J., and Ehrhardt, A. (2018) Novel vector construction based on alternative adenovirus types via homologous recombination, *Hum. Gene Ther. Methods 29*, 124–134.

[23] Zhang, W., Fu, J., Liu, J., Wang, H., Schiwon, M., Janz, S., Schaffarczyk, L., von der Goltz, L., Ehrke-Schulz, E., Dorner, J., Solanki, M., Boehme, P., Bergmann, T., Lieber, A., Lauber, C., Dahl, A., Petzold, A., Zhang, Y., Stewart, A. F., and Ehrhardt, A. (2017) An engineered virus library as a resource for the spectrum-wide exploration of virus and vector diversity, *Cell Rep. 19*, 1698–1709.

[24] Wang, J., Sarov, M., Rientjes, J., Fu, J., Hollak, H., Kranz, H., Xie, W., Stewart, A. F., and Zhang, Y. (2006) An improved recombineering approach by adding recA to λ red recombination, *Mol. Biotechnol. 32*, 43–54.

[25] Thomason, L. C., Sawitzke, J. A., Li, X., Costantino, N., and Court, D. L. (2014) Recombineering: genetic engineering in bacteria using homologous recombination, *Curr. Protoc. Mol. Biol. 106*, 1.16.11–11.16.39.

[26] Kolter, R., Inuzuka, M., and Helinski, D. R. (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K, *Cell 15*, 1199–1208.

[27] Thomason, L. C., Costantino, N., Shaw, D. V., and Court, D. L. (2007) Multicopy plasmid modification with phage λ Red recombineering, *Plasmid 58*, 148–158.

[28] Court, D. L., Sawitzke, J. A., and Thomason, L. C. (2002) Genetic engineering using homologous recombination, *Annu. Rev. Genet. 36*, 361–388.

[29] Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A. 97*, 5978–5983.

[30] Better, M., and Freifelder, D. (1983) Studies on the replication of *Escherichia coli* phage λ DNA I. The kinetics of DNA replication and requirements for the generation of rolling circles, *Virology 126*, 168–182.

[31] Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector, *Proc. Natl. Acad. Sci. U.S.A. 89*, 8794–8797.

[32] Niki, H., and Hiraga, S. (1997) Subcellular distribution of actively partitioning F plasmid during the cell division cycle in *E. coli*, *Cell 90*, 951–957.

[33] Gitai, Z. (2006) Plasmid segregation: a new class of cytoskeletal proteins emerges, *Curr. Biol. 16*, R133–R136.
[34] Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009) Recombineering: a homologous recombination-based method of genetic engineering, *Nat. Protoc. 4*, 206–223.
[35] Papa, L. J., III, and Shoulders, M. D. (2019) Genetic Engineering by DNA Recombineering, *Curr. Protoc. Chem. Biol. 11*, e70.
[36] Kaufmann, M. E. (1998) Pulsed-field gel electrophoresis, *Methods Mol. Med. 15*, 33–50.
[37] Bigot, S., Sivanathan, V., Possoz, C., Barre, F. X., and Cornet, F. (2007) FtsK, a literate chromosome segregation machine, *Mol. Microbiol. 64*, 1434–1441.
[38] Kuhlman, T. E., and Cox, E. C. (2010) Site-specific chromosomal integration of large synthetic constructs, *Nucleic Acids Res. 38*, e92.
[39] Hollenback, S. M., Lyman, S., and Cheng, J. (2011) Recombineering-based procedure for creating BAC transgene constructs for animals and cell lines, *Curr. Protoc. Mol. Biol. Chapter 23*, Unit23.14.
[40] Sawitzke, J. A., Thomason, L. C., Bubunenko, M., Li, X., Costantino, N., and Court, D. L. (2013) Recombineering: using drug cassettes to knock out genes *in vivo*, *Methods Enzymol. 533*, 79–102.
[41] Clarke, M. L., Burton, R. L., Hill, A. N., Litorja, M., Nahm, M. H., and Hwang, J. (2010) Low-cost, high-throughput, automated counting of bacterial colonies, *Cytom. A 77*, 790–797.
[42] Geissmann, Q. (2013) OpenCFU, a new free and open-source software to count cell colonies and other circular objects, *PLoS One 8*, e54072.

Chapter 3: Continuous Directed Evolution in Human Cells

Portions of the work presented in this chapter have been adapted from the following manuscript and appear in the thesis of Dr. Chet Berman:

Berman, C. M.*; Papa, L. J., III*; Hendel, S. J.*; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N.-D.; Kumar, C.; M., Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D. "An adaptable platform for directed evolution in human cells" *Journal of the American Chemical Society* **2018**, 140, 18093–18103.

*Denotes equal contributions

3.1 Author Contributions

L.J.P., C.M.B., and S.J.H. designed and performed experiments throughout the chapter. L.J.P. performed recombineering for the construction of the adenoviral vectors. C.M.B. and S.J.H. engineered cell lines and performed tissue culture experiments related to platform characterization. S.J.H. performed adenoviral protease trans-complementation experiments. S.J.H. performed the directed evolution of the tet-transactivator. S.J.H. and C.M.B. characterized tet-transactivator mutants. L.J.P. and C.M.B. constructed the selection circuits for Leucyl-tRNA synthetase and Cre recombinase. C.M.B. performed Leucyl-tRNA synthetase and Cre recombinase selection. L.J.P. and N.-D.D. synthesized the adenoviral protease inhibitor. C.L.M. was involved in project conception and data analysis. P.H.S. and A.F.W. performed experiments and analyzed data. C.M.K. assisted in cloning and recombineering. T.G.U. and R.C.H. designed the error-prone adenoviral polymerase and characterized its mutagenesis activity. Vincent L. Butty analyzed deep sequencing data. M.D.S. conceived the project, acquired funding, and supervised experiments and data analysis.

3.2 Introduction

Directed evolution methodologies have transformed our ability to generate biomolecules with improved or novel functionalities.[1-6] The vast majority of directed evolution experiments are performed in acellular environments, bacteria, or yeast. While these strategies have yielded many successes, they also frequently lead to products that fail to function optimally when later introduced into complex metazoan systems. The evolved functions can be derailed by off-target interactions, poor protein folding or stability, pleiotropic outputs, or other serious problems that arise because the biomolecules were discovered and optimized in overly simplistic environments.[7-9] This frontier challenge could be most directly addressed by leveraging the human cell itself as the design, engineering, and quality control factory for directed evolution-mediated biomolecule discovery and optimization.

Extant strategies for directed evolution in human cells rely almost entirely on phenotypic or fluorescent screens to identify active biomolecule variants. The most common technique is *in vitro* plasmid mutagenesis followed by transfection and screening.[10] This approach is slow, labor-intensive, and significantly constrains library size. Other methods include *in vivo* mutagenesis through somatic hypermutation in immune cells followed by screening or selection[11,12] More recently, robotic cell-picking techniques have been used to comprehensively screen for desired phenotypes across multiple dimensions (e.g., both extent and localization of a fluorescent signal).[9] These methods, while valuable, are still slow and inefficient and have limited library sizes (~10^5 variants for the most recent robotic platform).[9] Another recent development has been the use of cytidine deaminase fused to Cas9 variants to introduce mutations into endogenous genes in human cells, followed by selecting or screening for desired phenotypes.[13-15] However, these methods require the design and synthesis of many guide RNAs to tile along regions of interest, which requires repeated rounds of sequencing and guide RNA redesign as mutations accumulate. Moreover, directed evolution achieved via *in vivo* mutagenesis of the human genome is limited by the slow growth rate of human cells and the high potential for false positives ("cheaters") associated with any strategy that relies on cell selection or screening.

A broadly useful human cell-based directed evolution platform requires several critical features: (1) Large mutational libraries expressed in the human cell; (2) Selection schemes providing a broad dynamic range for selection and minimal opportunities for cheating; (3) Capacity to evolve multiple biomolecule functions; (4) Applicability across multiple cell types; and (5) Ideally, a minimal need for experimenter intervention during evolution experiments.

Inspiration for such a platform can be drawn from prior efforts coupling biomolecule function to viral replication using HIV[16] or bacteriophage.[17] However, HIV-based strategies suffer from an inability of the virus to propagate under strong selection pressure or in most cell types, and raise safety concerns surrounding large-scale HIV culture. The M13 bacteriophage used in phage-assisted continuous evolution[17] provides large mutational libraries and enables rapid rounds of selection and mutagenesis for biomolecules carrying out diverse functions, but only permits directed evolution in bacterial cells.

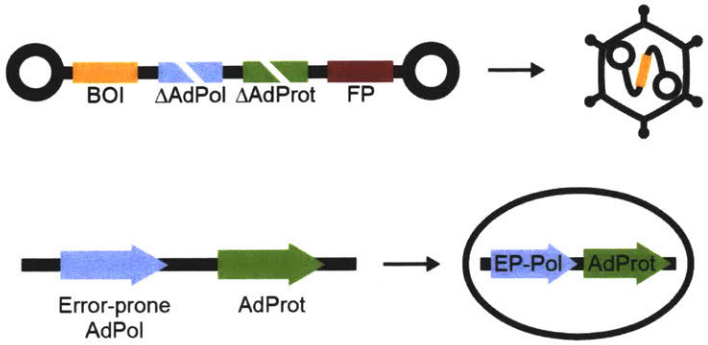

Figure 3.1 Human cell-based directed evolution components.
Schematic of an engineered adenovirus type-5 vector (*top*) in which genes for adenoviral polymerase (AdPol) and protease (AdProt) are removed and a gene encoding the biomolecule of interest (BOI) for directed evolution is introduced, as well as a fluorescent protein (FP) for visualization during infection. Engineered human cells constitutively express a highly error-prone AdPol (termed EP-Pol) and conditionally expressing AdProt at levels directly dependent on BOI activity (*bottom*).

With these parameters and challenges in mind, we aimed to devise a broadly useful human cell-based directed evolution platform. We rationalized that adenovirus type-5 would be a practical vector for the directed evolution of biomolecules in human cells, owing to its genetic tractability and broadly infectious nature in many human cell types. Furthermore, decades of research have shown that adenovirus tolerates an extremely wide range of transgenes, ensuring broad applicability of an adenovirus-based platform to diverse directed evolution targets. Conceptually, if the replication of a highly mutagenic adenovirus somehow depended on the activity of a biomolecule of interest (BOI) encoded in the adenoviral genome, then a simple directed evolution scheme for evolving diverse BOI functions in human cells could be feasible.

To achieve this concept, we first deleted the essential adenoviral DNA polymerase (AdPol) and protease (AdProt) genes from an adenoviral genome that also encoded the BOI for evolution (**Figure 3.1**). The resulting adenovirus deletion variant is incapable of replication outside engineered human cells. We trans-complemented the missing AdPol by constitutive expression, within human cells, of a newly designed, highly mutagenic AdPol variant to enable the generation of large mutational libraries during viral replication. AdProt expression in the human cells was then engineered to depend conditionally upon BOI function (**Figure 3.1**). Directed evolution experiments in this system relies on simply serially passaging the BOI-encoding adenovirus, with mutagenesis and selection continuously occurring (**Figure 3.2**).

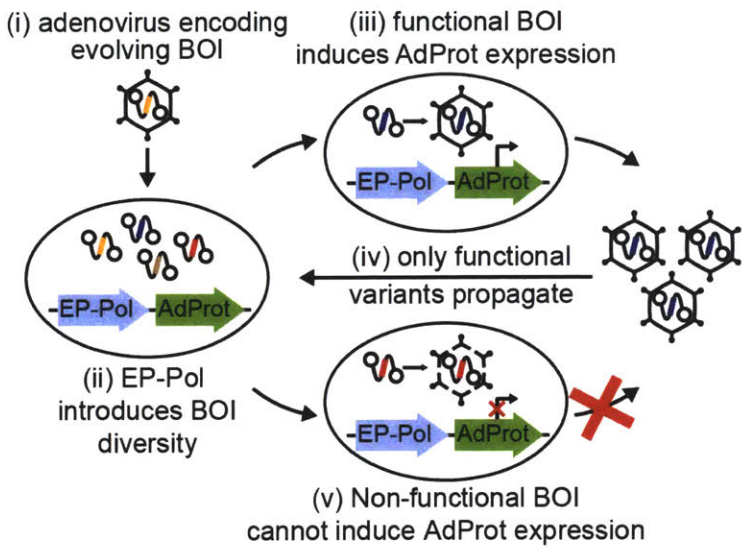

Figure 3.2 Schematic for adenovirus-based directed evolution of BOIs in human cells.
(i) The BOI is delivered into the human cell via adenoviral infection. (ii) EP-Pol introduces mutations into the BOI gene, generating a mutational library. (iii) The desired BOI function is coupled to the expression or activity of AdProt such that (iv) only functional BOI variants result in viral propagation. (v) If the BOI variant is non-functional, AdProt is not expressed or active and the adenovirus encoding that variant is outcompeted.

Here, we present the key features of this new platform, including mutagenesis, selection, and enrichment parameters. We further demonstrate the platform's utility via proof-of-concept directed evolution experiments in which we evolved, directly in the human cell environment, multiple transcription factor variants that maintained high levels of function while gaining resistance to a small-molecule inhibitor. Altogether, we believe that this platform holds significant potential to not only enable the development of new research tools, but also enhance our understanding of metazoan evolutionary biology and our ability to rapidly generate and optimize biomolecular therapeutics.

3.3 Results

3.3.1 Mutagenesis

Adenovirus type-5 relies on its own DNA polymerase, AdPol, for replication of its double-stranded DNA genome.[18] The high fidelity AdPol has an estimated mutation rate of ~1.3 × 10^{-7} mutations per base per viral passage, based on high fidelity deep sequencing experiments performed by Sanjúan and co-workers.[19] Such a low mutation rate is insufficient to generate the large library sizes necessary for laboratory time-scale directed evolution. We therefore sought to increase the mutation rate of adenovirus by engineering a highly mutagenic variant of AdPol.

Previous studies identified two amino acid substitutions in AdPol, F421Y and D827A, that separately increase the mutation rate of AdPol, likely through distinct mechanisms (**Figure 3.3a**).[20] In the φ29 bacteriophage polymerase,[21] an AdPol homolog, the amino acid analogous to F421 occurs in the proofreading exonuclease domain, suggesting that the F421Y AdPol variant may have weakened proofreading capacity. The amino acid analogous to D827 occurs in the fingers domain involved in selection of incoming nucleotides, again suggesting a possible mechanism for the reduced fidelity of D827A AdPol. We reasoned that combining these two substitutions to create the F421Y/D827A AdPol double-mutant, which we termed error-prone AdPol (or EP-Pol), would allow us to further increase the mutation rate while still supporting robust adenovirus propagation.

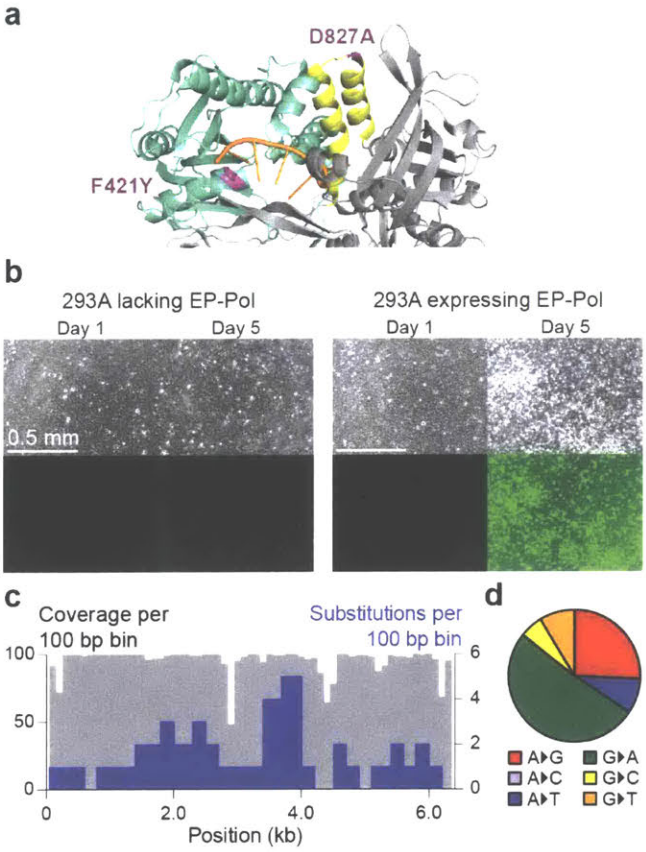

Figure 3.3 Error prone polymerase introduces mutations into ΔAdPol adenovirus.
(**a**) Crystal structure of the φ29 DNA polymerase (PDBID 1XHZ)[21] an AdPol homolog, with the locations of homologous mutations used to create EP-Pol shown in magenta. (**b**) Either parental HEK293A cells or cells constitutively expressing EP-Pol were infected with a GFP-encoding ΔAdPol-adenovirus (CFP.ΔAdPol.GFP). The virus propagated only on EP-Pol trans-complementing cells. Similar results were obtained for wild-type AdPol (**Figure 3.4**). (**c**) ΔAdPol-adenovirus (AdGLΔPol) was serially passaged on EP-Pol expressing cells for ten passages, after which a 6.5 kb genomic fragment was amplified from an ~27 clone pool. Illumina sequencing identified mutations throughout the amplified region. For substitution values, see **Table 3.3**. (**d**) Mutational spectrum of EP-Pol evaluated by next-generation sequencing.

To test this hypothesis, we first used recombineering to inactivate the AdPol gene encoded by the adenovirus type-5 genome via an internal deletion (see **Table 3.1** for a list of adenoviral constructs employed). Next, we stably transduced HEK293A cells with an HA-tagged version of either wild-type AdPol or EP-Pol (see **Table 3.2** for a list of cell lines employed). We observed that ΔAdPol adenoviruses (CFP.ΔAdPol.GFP where CFP and GFP correspond to cyan and green fluorescent protein, respectively) propagated only

on cells that expressed either AdPol (**Figure 3.4**) or EP-Pol in *trans* (**Figure 3.3b**). Further, we observed that EP-Pol and wild-type AdPol both supported robust ΔAdPol-adenovirus replication.

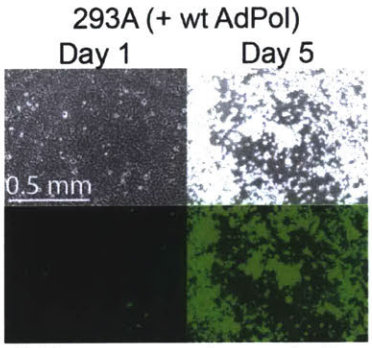

Figure 3.4 Trans-complementation of wild-type adenoviral polymerase (AdPol).
Parental HEK293A cells stably expressing wt AdPol were infected with a GFP-encoding ΔAdPol-adenovirus (CFP. ΔAdPol.GFP). The virus propagated robustly on these AdPol expressing cells.

We next assessed the mutation rate endowed by EP-Pol. After passaging ΔAdPol-adenovirus (AdGLΔPol) on EP-Pol trans-complementing human cells for 10 serial passages, we deep sequenced a 6.5 kb region of the genome obtained from a pool of about 27 viral clones (**Figure 3.3c**; see also **Table 3.3**). This sequencing revealed a mutation rate of 3.7×10^{-5} mutations per base per passage. As the adenoviral genome is ~36 kb, this mutation rate indicates that EP-Pol introduced ~1.3 mutations into the genome per infected cell per passage. Moreover, EP-Pol displayed a broad mutational spectrum, including both transitions and transversions (**Figure 3.3d**).

Table 3.1 Adenoviruses constructed and used in this study.

Name	Modifications relative to wild-type Ad5
AdCFP	E1R-CFP ΔE1 ΔE3
CFP.ΔAdPol.GFP	E1R-CFP ΔE1 ΔE3 ΔAdPol E4R-GFP
tTA_{wt}.mCherry	E1L-tTA ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
tTA_{mut}.GFP	E1L-tTA_{aak} ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-GFP
AdEvolve-DEST	E1L-DEST ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
CFP.ΔAdProtΔAdPol	E1R-CFP ΔE1 ΔE3 ΔAdProt ΔAdPol
AdGLΔPol[20]	E1L-Luciferase-GFP ΔE1 ΔE3
Cre.Ad	E1L-Cre ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
LeuRS.Ad	E1L-LeuRS ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry

Note: All viruses used in this work were derived from AdCFP except for AdGLΔPol, which was previously reported.[20]

Table 3.2 Cell lines used in this study.

Cell line	Polymerase	Transgene cassette
Producer	AdPol	CMV.AdProt
Mutator	EP-Pol	CMV.AdProt
Selector	EP-Pol	TRE3G.AdProt
Phenotyping	AdPol	TRE3G.eGFP

Note: All cell lines were derived from HEK293A cells.

Table 3.3 Tabulation of next-generation sequencing results and experimental parameters used to estimate the EP-Pol mutation rate.[20]

Estimated number of clones sequenced	Size of the region sequenced and analyzed (bp)	Substitution load per million bp	Substitutions per Ad genome per viral generation
27.3*	6020	365	1.31**

* Viral pool size was estimated based on intra-experiment titrations during pool preparations

**Assuming a genome size of 36 kb and that 27.3 genomes were sequenced. Each of the 10 passages was defined as a generation.

Previously, the same sequencing procedure was carried out for wild-type AdPol.[20] Because only one mutation introduced by wild-type AdPol was detected across two separate trials in that experiment, it was not possible to define an actual mutation rate for wild-type AdPol. In contrast, 60 mutations and 13 insertions were observed for EP-Pol. Compared to the previously reported mutation rate of wild-type AdPol determined by another method,[19] however, the mutation rate of EP-Pol is enhanced ~280-fold. Thus, EP-Pol greatly increases the number of mutations introduced per viral passage. Based on these analyses, the EP-Pol mutation rate is similar to highly mutagenic RNA viruses that can readily evolve on laboratory timescales.[22-24]

We next estimated the lower limit of the library size in a given passage (or 'round') of directed evolution using EP-Pol. A typical round of directed evolution might reasonably involve infecting 3.0×10^8 human cells at a low MOI. Each round of directed evolution could conclude once 100% of cells (~3.0×10^8 cells) are infected. Because ~1.3 mutations are introduced per cell per replication, and because there is at least one replication in each round of evolution since the infection occurs at low MOI, we estimate that there are ~4×10^8 adenoviral variants after one passage. Assuming a typical 1 kb gene encoding the BOI composes ~1/30 of the engineered adenoviral genome, there would be ~1.3×10^7 variants of the BOI in the population after one round of evolution. This calculation is a lower limit because it does not account for any genetic diversity at the beginning of each round. Regardless, even this conservative estimate indicates that we can generate virtually all single, many double, and some triple mutants in a typical BOI in a single round of evolution. Notably, the mutations are continuously introduced instead of requiring in vitro mutagenesis physically separated from selection and propagation steps.

3.3.2 Selection

Our next objective was to design an appropriate selection scheme capable of coupling BOI activity to adenoviral propagation. After extensive testing of assorted adenoviral genes, we developed such a scheme based on deleting the gene for adenoviral protease (AdProt) from the viral genome and then providing AdProt in trans from the human host cell.[25] AdProt has vital functions in viral uncoating, DNA replication, viral maturation, and cell entry.[26,27] Importantly, AdProt is a 'late gene' expressed mainly after DNA replication of the adenoviral genome.[27] Because AdProt is not required in the early stages of infection, BOI variants can be generated by mutagenesis before selection pressure is applied during a given infection.

We began by testing whether AdProt trans-complementation could be achieved in the context of an adenovirus already requiring AdPol trans-complementation. We stably expressed AdProt in an AdPol-expressing cell line, termed "producer" cells (see **Table 3.2**). Next, we monitored the progress of an adenovirus infection of ΔAdProtΔAdPol-adenovirus on AdPol-expressing versus AdPol- and AdProt-expressing cells. We observed that only the cell line constitutively expressing both AdProt and AdPol supported robust replication of ΔAdProtΔAdPol-adenovirus (**Figure 3.5**). Thus, host cell expression of AdPol and AdProt can successfully support the replication of an AdPol- and AdProt-deleted adenovirus, permitting both the facile production of ΔAdProtΔAdPol-adenoviruses and providing a potential mechanism to impart selection pressure in a directed evolution experiment.

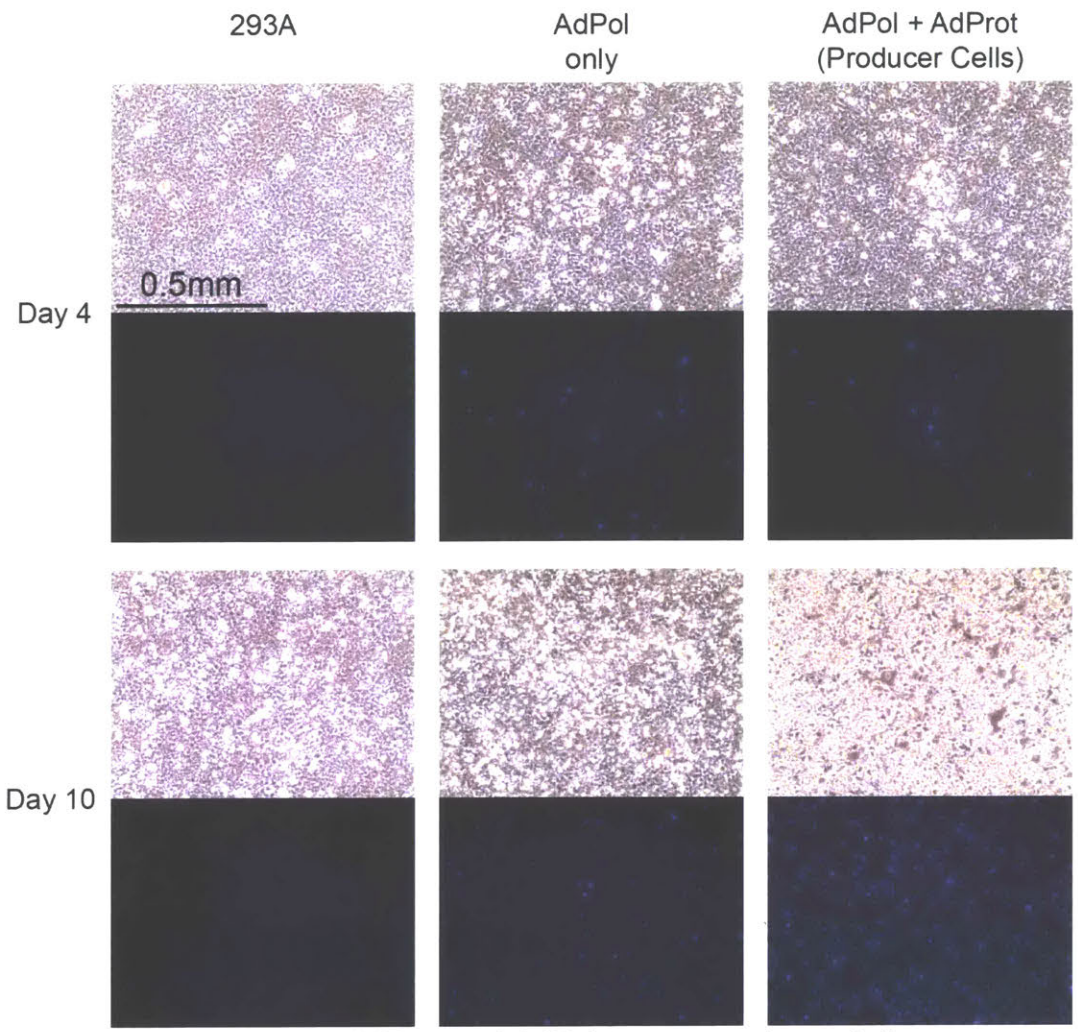

Figure 3.5 Simultaneous AdProt and AdPol trans-complementation.
CFP.ΔAdProtΔAdPol was used to infect HEK293A cells, AdPol-expressing cells, or producer cells (**Table S2**) at a low multiplicity of infection (< 0.5). The infection was monitored over 10 days. The parental HEK293A cells showed no visible sign of infection, likely because without AdPol expression the copy number of the CFP gene was too low in the cell to easily visualize fluorescence with laboratory microscopes. The AdPol-expressing cells showed a strong CFP signal indicating a robust infection, however the infection did not spread owing to a lack of AdProt. In contrast, the producer cells trans-complementing both AdPol and AdProt were able to support a spreading infection, with every cell in the plate infected by day 10.

We next evaluated the capacity of this AdProt-complementation strategy to confer sufficient selection pressure to drive a directed evolution workflow. For this purpose, we performed a competition experiment on a model BOI, tTA.[28,29] Wild-type tTA (tTA_{wt}) binds its endogenous operator, with a consensus sequence of 5′-CCTATCAGTGATAGA-3′, to induce downstream gene transcription. A tTA variant (tTA_{mut}) that is incapable of binding to the endogenous operators has also been reported.[30] tTA_{mut} instead possesses enhanced affinity for the mutant 5′-CCcgTCAGTGAcgGA-3′ operator. We engineered ΔAdProtΔAdPol-adenoviruses that expressed either tTA_{wt} and mCherry (tTA_{wt}.mCherry)

or tTA_{mut} and GFP (tTA_{mut}.GFP). We then stably transduced AdPol-expressing HEK293A cells with a lentiviral vector that provided AdProt under control of the endogenous tTA operator (termed "selector" cells, see **Table 3.2**). In this cell line, tTA_{wt}.mCherry adenovirus should be able to strongly induce AdProt and propagate, whereas tTA_{mut}.GFP should not induce AdProt and therefore should not form infectious virions. Because these viruses express different fluorescent markers, relative viral populations can be assessed using flow cytometry upon infection of human cells that do not express AdProt in order to prevent propagation and therefore more accurately quantify the resulting viral populations.

To test our hypothesis that AdProt induction could enable enrichment of active over inactive BOI variants, we co-infected tTA_{wt}.mCherry and tTA_{mut}.GFP using a total MOI of ~0.25 in selector cells (see **Table 3.2)** at initial ratios of 1:100 or 1:1,000 (**Figure 3.6a**). We then performed three serial passages on selector cells, and analyzed the resulting viral populations via infection of AdPol-expressing but AdProt-lacking HEK293A cells followed by flow cytometry. In the initial passage, the tTA_{wt}.mCherry adenovirus enriched at least 40–50-fold over the tTA_{mut}.GFP adenovirus (**Figure 3.6b**). Furthermore, across three rounds of passaging, the tTA_{wt}.mCherry adenoviruses were consistently enriched to > 90% of the adenoviral population regardless of the starting ratios. Thus, our AdProt-based selection strategy can rapidly enrich active BOIs that are initially present at low frequency in a viral population.

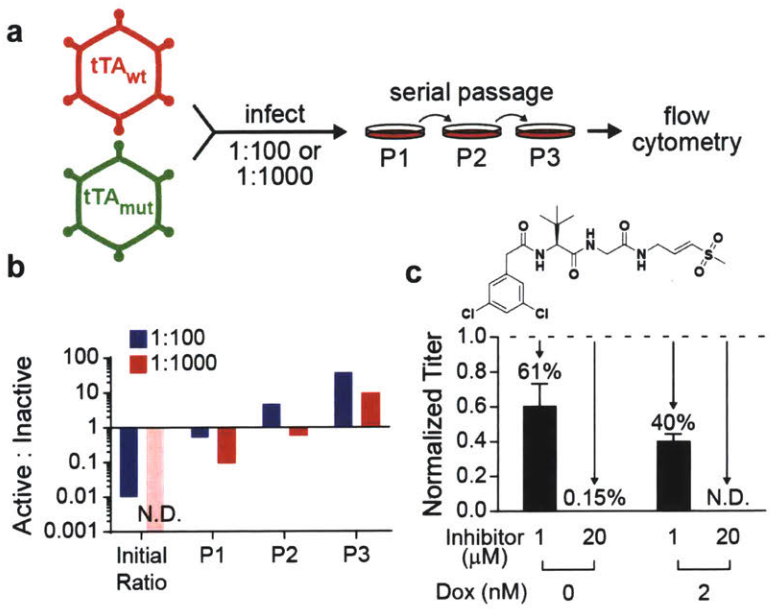

Figure 3.6 Enrichment of active tTA-carrying adenovirus and tuning selection pressure with a small molecule inhibitor.

(**a**) Schematic of the competition experiment between adenoviruses that carry the gene for wild-type tetracycline transactivator (tTA$_{wt}$.mCherry) versus viruses that carry inactive tTA (tTA$_{mut}$.GFP). HEK293A cells stably encoding the gene for adenoviral protease (AdProt) under control of the endogenous tTA operator are infected by an initial ratio of 1:100 or 1:1,000 tTA$_{wt}$.mCherry to tTA$_{mut}$.GFP viruses. Viral media was serially passaged onto a new plate of cells for three rounds. The viral populations were then determined via flow cytometry. (**b**) Quantification of flow cytometry data from the competition experiment. The proportion of tTA$_{wt}$.mCherry adenoviruses relative to tTA$_{mut}$.GFP adenoviruses rapidly increased with each passage. The initial ratio of the 1:1,000 sample (labeled N.D.; not detectable) was not experimentally quantifiable owing to the low amount of tTA$_{wt}$.mCherry adenovirus present, and was therefore derived by dilution of the 1:100 initial ratio. For raw flow cytometry data, see **Figure 3.7** and **Figure 3.8**. (**c**) AdProt-based selection pressure in combination with administration of a small molecule AdProt inhibitor (structure shown) provides access to an orders of magnitude-wide dynamic range of selection pressure. tTA-inducible AdProt cells were infected with tTA$_{wt}$.mCherry adenovirus, and treated with a combination of doxycycline (dox) and the AdProt inhibitor. The resulting viral supernatant was titered by flow cytometry. Titers were normalized to infections performed in the absence of the AdProt inhibitor. The titer of the adenovirus treated with 20 μM AdProt inhibitor and 2 nM dox was too low to be accurately detected (N.D.; not detectable).

We next applied this tTA-based genetic circuit to evaluate the dynamic range of AdProt selection. Our approach was to employ an allosteric inhibitor of tTA, doxycycline (dox), to tune AdProt expression levels. In the presence of dox, tTA is unable to bind its target operator and induce AdProt expression. Using this approach, based on AdProt transcript levels we were able to access up to a 14-fold change in AdProt expression

(**Figure 3.9a**). Notably, we observed a strong correlation between dox concentration and viral titer over this entire order of magnitude range (**Figure 3.9b**).

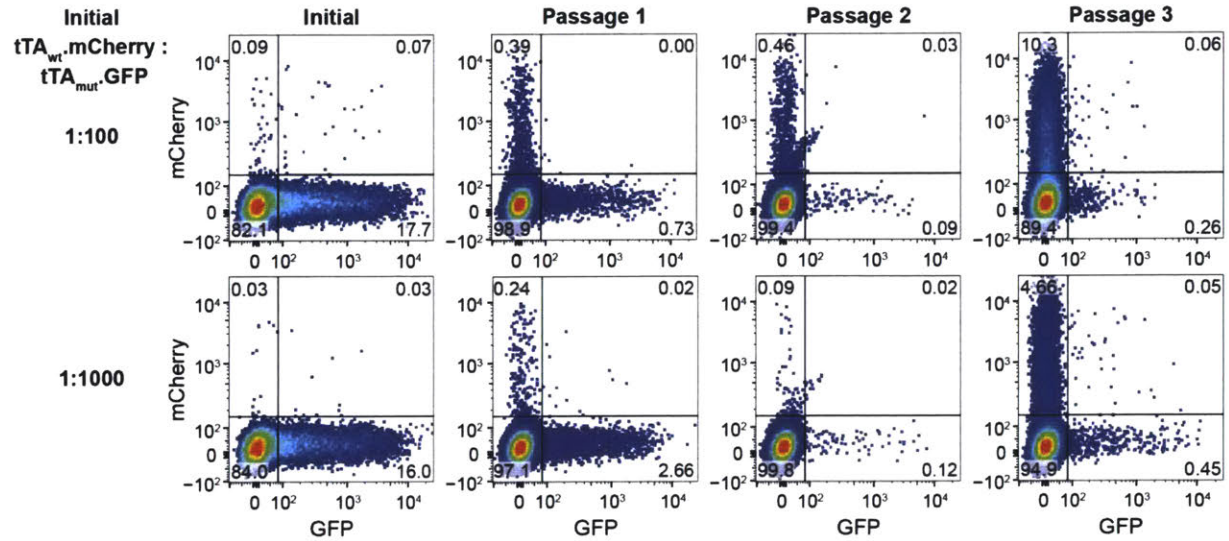

Figure 3.7 Enrichment of tTA$_{wt}$.mCherry over tTA$_{mut}$.GFP adenovirus.
Flow cytometry data showing infection with mixed samples of tTA$_{wt}$.mCherry adenoviruses and tTA$_{mut}$.GFP adenoviruses (Table 3.1
Table 3.1) over three serial passages. Density plots show cells infected with tTA$_{wt}$.mCherry adenoviruses (Q1, upper left), tTA$_{mut}$.GFP adenoviruses (Q4, lower right), or both (Q2, upper right). Quantifications of each quadrant as a percentage of the total population are shown.

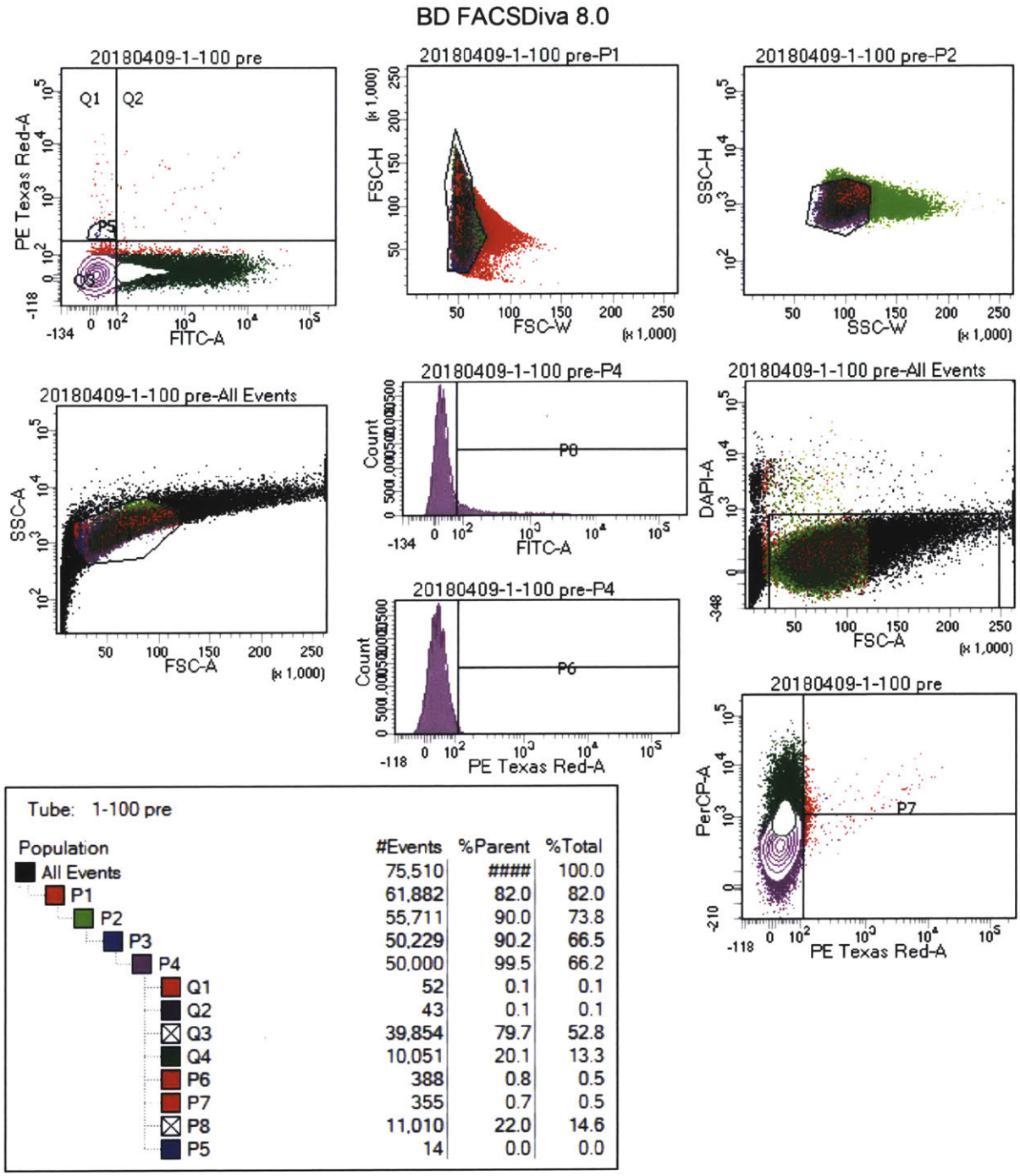

Figure 3.8 Example raw flow cytometry data.
Gates P1–P3 were used to eliminate cell debris and cell aggregates. Gate P4 was used to remove dead cells by only gating for DAPI-negative cells. Gates Q1–Q4 were used to gate for GFP-positive and mCherry-positive cells. This specific data set was used to calculate the initial ratio of tTA_{wt}.mCherry virus to tTA_{mut}.GFP virus in the competition experiment (**Figure 3.6a** and **Figure 3.7**).

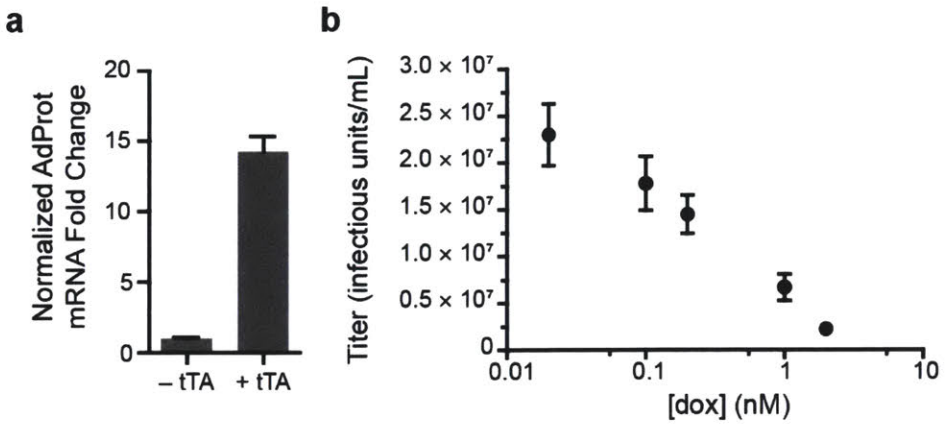

Figure 3.9 AdProt levels can be modulated by tTA and dox.
(a) AdProt transcript levels were analyzed by qPCR in selector cells 2 days after transfection with pTet-Off Advanced. AdProt transcript levels normalized to untransfected selector cells (**Table 3.2**). (b) Titer of tTA_{wt}.mCherry adenovirus after infection of selector cells (**Table 3.2**) treated with varying amounts of doxycycline (dox). Titers are reported as "infectious units/mL" defined as the number of fluorescent cells per mL of viral supernatant used during the infection.

We note that there is likely to be an upper bound to the number of active AdProt molecules required for replication, at which point additional AdProt induction will not result in greater viral replication. As a result, the selective advantage would be low or non-existent for any evolved BOIs that are able to induce AdProt above the upper bound, despite our desire to enrich such variants. A small molecule inhibitor of AdProt could provide a way to dynamically tune selection pressure by reducing AdProt activity per given number of AdProt molecules such that the upper limit of useful AdProt molecules is raised, thus conferring a selective advantage to highly active BOIs that would otherwise produce more AdProt molecules than necessary. The ability to raise this upper limit could be critical for maintaining selection pressure as a directed evolution experiment progresses and the average BOI variant in the pool becomes more and more active. Indeed, when we challenged tTA_{wt}.mCherry-expressing adenoviruses with various concentrations of the vinyl sulfone AdProt inhibitor shown in **Figure 3.6c**,[31] we found that the inhibitor could reduce the infectious titer of the tTA_{wt}.mCherry virus up to 650-fold, providing ready access to a dynamic range of selection pressure between 2–3 orders of magnitude in size. Moreover, we observed that the AdProt inhibitor even further reduced infectious titer in the presence of dox (**Figure 3.6c**), highlighting the capacity of AdProt inhibition to strengthen selection pressure at a variety of baseline AdProt expression levels. Notably, the vinyl sulfone AdProt inhibitor was not toxic at the concentrations used (**Figure 3.10**).

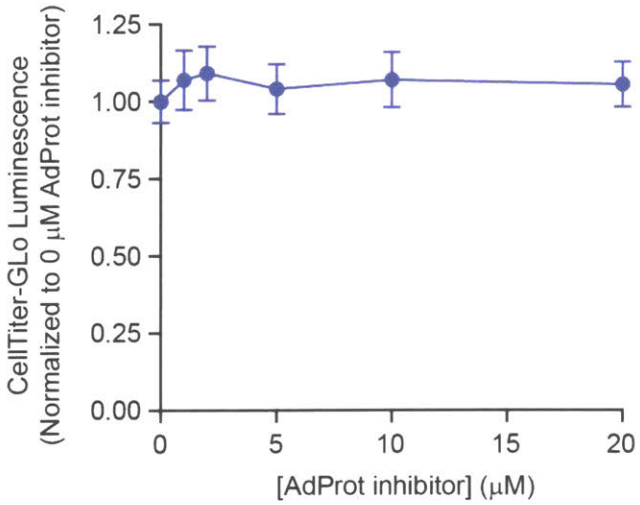

Figure 3.10 AdProt inhibitor concentrations used are not cytotoxic.

CellTiter-GLo assay to assess any impact of the AdProt inhibitor on cell viability at concentrations up to 20 μM (*N* = 5, 5-day treatment).

3.3.3 Directed evolution of functional, drug-resistant tTA variants in human cells

As a proof-of-concept, we aimed to evolve tTA variants that retained transcription-inducing activity but gained resistance to their small molecule inhibitor, dox. Specifically, we serially passaged our tTA_{wt}.mCherry virus in the presence of dox in a "selector" cell line (see **Table 3.2**) that inducibly expressed AdProt under control of the endogenous tTA operator. We maintained a low initial multiplicity of infection (~0.05) to minimize the probability that viruses encoding distinct tTA variants would co-infect the same cell, at least at an early stage of each passage. Co-infections could result in "hitchhiking," in which low fitness variants can be temporarily maintained in the population by infecting the same cell as high fitness variants. Such hitchhikers could slow the pace of selection. We transferred viral supernatant to fresh cell plates upon the appearance of spreading infection, with the goal of selecting for viruses that encode functional, but dox-resistant, tTA variants.

We ran two evolution experiments in parallel (Trials 1 and 2) with different selection pressure strategies (**Figure 3.11a**). In Trial 1, we tuned the selection pressure over time, increasing the dox concentration from 2 nM up to 20 μM. In Trial 2, we kept selection pressure constant and high by maintaining the dox concentration at 200 nM. In order to test whether dox-resistant tTA variants enriched in the population, we used the viral media from each passage in Trial 1 to infect a "phenotyping" cell line (see **Table 3.2**) containing GFP under control of the endogenous tTA operator in the presence of dox. This phenotyping cell line lacked AdProt, allowing the virus to infect the cells and induce GFP expression, but not to proliferate. We measured GFP induction by the viral population harvested after each serial passage in the presence of 20 μM dox in these phenotyping cells using flow cytometry (**Figure 3.11b**). Substantial dox-resistant tTA activity emerged by passage 5, suggesting that dox-resistant variant(s) of tTA may have arisen and enriched in the viral population.

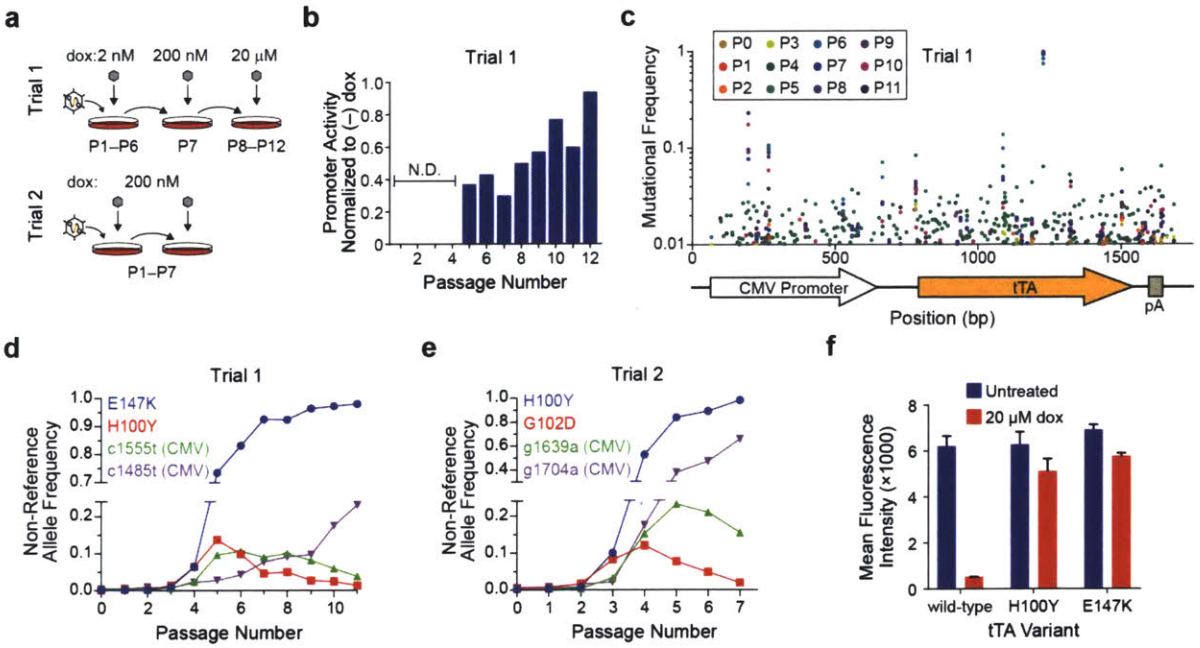

Figure 3.11 Passaging tTA-carrying adenovirus in the presence of dox evolves dox-insensitive tTA variants.

(**a**) Serial-passaging schemes for evolving functional tTA variants that gain dox resistance in human cells. Two approaches to selection pressure were used, either with increasing dox concentrations (Trial 1) or a constant, moderate dox concentration (Trial 2). (**b**) tTA-induced GFP expression in the presence of dox after each round of evolution for Trial 1. Phenotyping cells were infected with passaged viral populations and analyzed by flow cytometry. The percentage of infected GFP-positive cells at each passage in the presence of dox was normalized to the percentage of infected GFP-positive cells at each passage in the absence of dox. N.D. = not detectable owing to low viral titer. (**c**) Non-reference allele frequencies for all mutations observed at ≥ 1% frequency over the course of the directed evolution experiment for Trial 1 (see **Figure 3.12** for Trial 2). A schematic of the sequenced amplicon is shown below the x-axis for reference. (**d**) Mutational trajectories of four mutations identified in Trial 1, including two non-coding mutations in the CMV promoter upstream of the tTA gene. (**e**) Mutational trajectories of four abundant mutations identified in Trial 2, including two non-coding mutations in the CMV promoter upstream of the tTA gene. (**f**) Plasmids encoding the tTA variants that fixed in Trials 1 and 2 were transfected, along with the pLVX-TRE3G.eGFP reporter plasmid, into HEK293A cells with or without dox (N = 3). Two days later, flow cytometry was performed to examine tTA variant activity in the presence versus the absence of 20 μM dox.

We next examined whether mutations in the tTA gene contributed to this decreased dox sensitivity. We amplified and sequenced a 1.75-kb region of the adenoviral genome containing the tTA open reading frame from virus harvested at each passage during both Trials. Using this approach, we detected > 200 unique mutations that attained ≥ 1% frequency by passage 4 in Trial 1 (**Figure 3.11c**), even though promoter activity at passage 4 was still undetectable (**Figure 3.11b**). In Trial 2, 43 mutations attained ≥ 1% by passage 4 (**Figure 3.12**). By passage 5, a single amino acid substitution in tTA attained > 70% frequency in the viral population in both trials (E147K in Trial 1 and H100Y in Trial 2), rapidly becoming fully fixed in the population thereafter (**Figure 3.11d** and **Figure**

3.11e). Both mutations observed were previously reported to confer dox-resistance in tTA,[32] which we further confirmed through transient co-transfection of a plasmid encoding GFP under control of the endogenous tTA operator along with wild-type, E147K, or H100Y tTA-encoding plasmids into HEK293A cells in the presence or absence of dox (**Figure 3.11f**). Additional mutations that were also previously reported[32] to confer dox-resistance were also observed at > 10% frequency early in the directed evolution experiment (H100Y in Trial 1 and G102D in Trial 2).

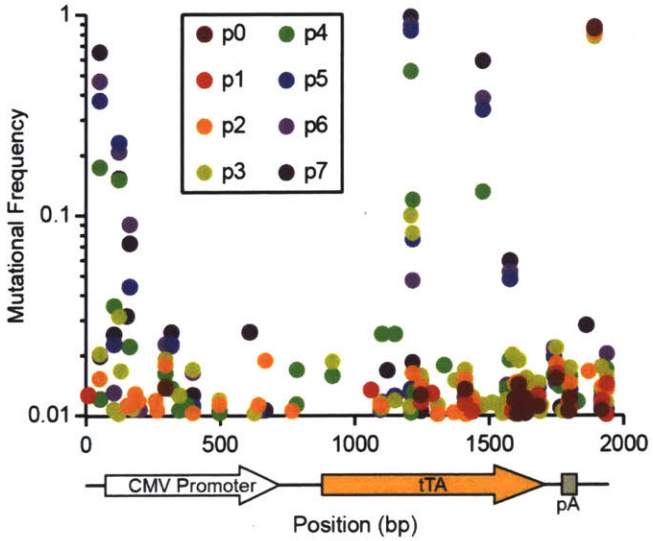

Figure 3.12 Allele frequencies at each passage. Non-reference allele frequencies at ≥ 1% frequency over the course of the directed evolution experiment for Trial 2.

In Trial 2, we also analyzed the possible effects of hitchhikers on the enrichment of active variants. Our approach was to harvest the adenovirus at two different timepoints: (i) either early, when ~75% of cells were infected and co-infection was minimized or (ii) very late, after full cytopathic effect was achieved and most cells were likely to be co-infected. We found that even under very high co-infection conditions (late harvest), variants previously reported to be dox-resistant[32] continued to enrich, possibly even more than under low co-infection conditions (**Figure 3.13**). Thus, co-infection did not hinder the enrichment of active variants.

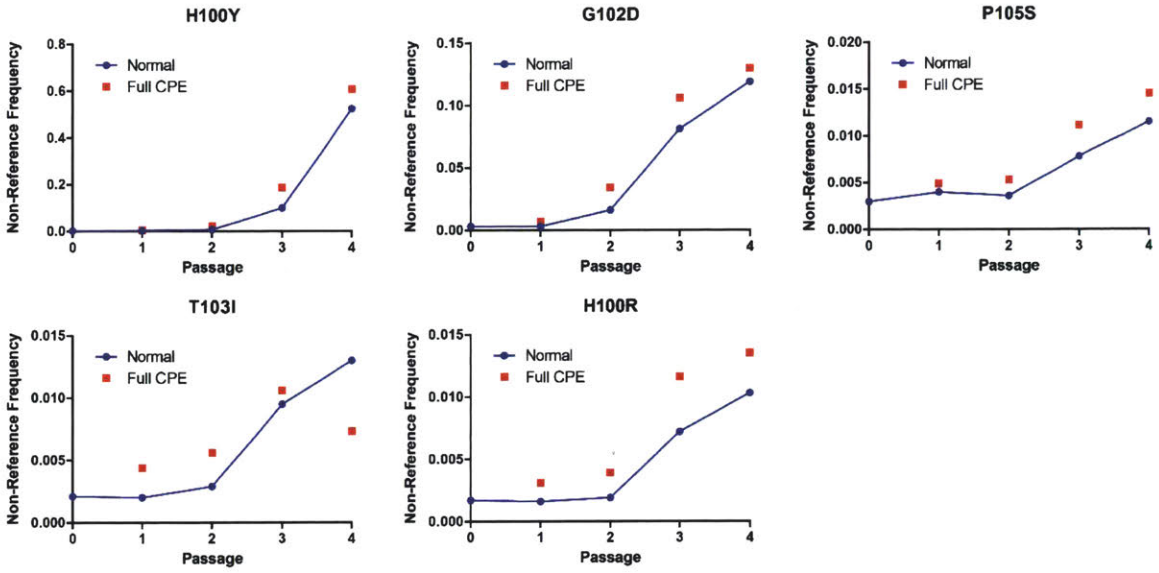

Figure 3.13 Frequencies of previously reported dox-resistant alleles.
Comparison of mutation frequencies in Trial 2 using the early harvesting protocol (75% of cells infected) and late harvesting for each passage after full CPE was attained. Five previously reported doxycycline-resistant variants that reached a frequency of 1% by passage 4 are shown.

These results highlight both the different outcomes that can result from repeated evolution experiments and the capacity of our platform to explore sequence space in human cells. Additionally, we were able to evolve biomolecules using two different selection pressure protocols (gradually increasing pressure or constant, high pressure). In summary, our directed evolution protocol can successfully generate and rapidly enrich functional BOI variants in human cells, merely by serial passaging of a BOI-encoding adenovirus.

3.3.4 Design of alternative selection circuits

In the interest of highlighting the utility of our platform beyond the directed evolution of transcription factors, we sought to demonstrate how alternative selection circuits could be used to evolve different types of functions. We created two new selection circuits for a user-defined recombinase activity and aminoacyl-tRNA synthetase activity (**Figure 3.14**).[33,34] We transfected both the Cre-recombinase (Cre, **Figure 3.14a**) and leucyl-tRNA synthetase (LeuRS, **Figure 3.14b**) AdProt selection circuits into HEK293A cells expressing AdPol and then monitored the replication of ΔAdProtΔAdPol-adenoviruses expressing Cre, LeuRS, or a control, inactive BOI (tTA). For the recombinase circuit, we found that the Cre-containing adenovirus replicated >20-fold better than a control adenovirus (**Figure 3.14c**). For the aminoacyl-tRNA synthetase circuit, we observed the LeuRS-containing adenovirus was able to replicate while the control adenovirus could not replicate to detectable levels (**Figure 3.14c**). All adenoviruses replicated robustly on a control circuit that constitutively expressed protease. These data indicate that our platform can be easily adapted to select for desired recombinase and aminoacyl-tRNA synthetase activities, and shows that selection circuits can be introduced through transient transfection as well as stable integration.

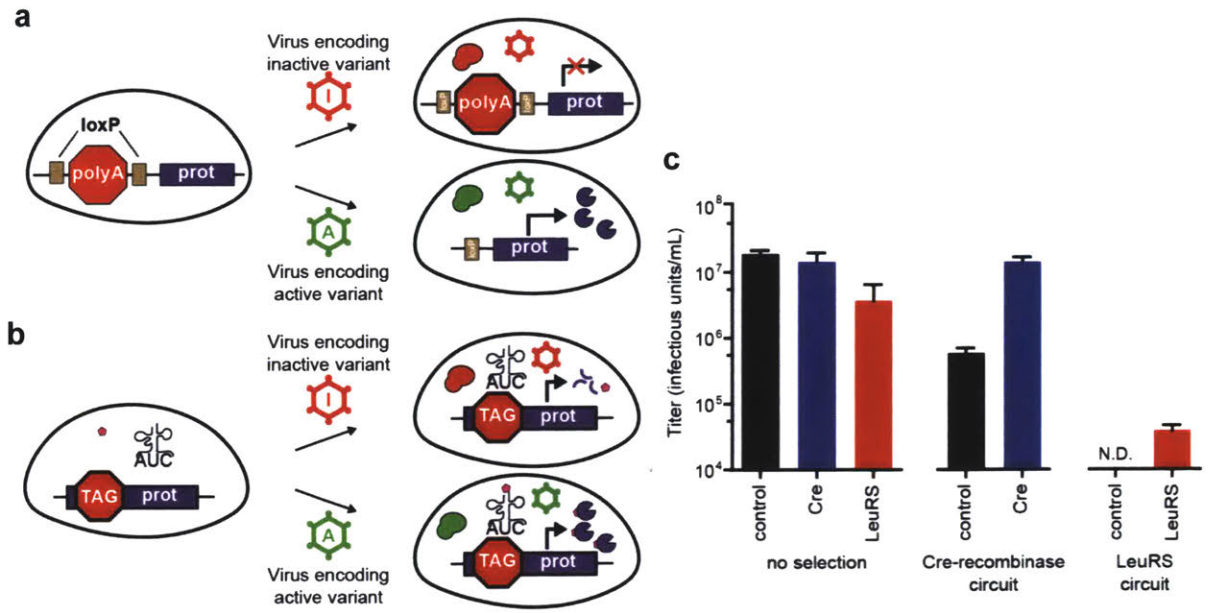

Figure 3.14 Testing of Cre recombinase- and leucyl-tRNA synthetase-based selection circuits.

(**a**) Selection circuit designed for AdProt-based selection of Cre recombinase activity. A floxed SV40-polyA terminator signal prevents transcription of AdProt unless Cre deletes the terminator by recombination at target loxP sites. (**b**) Selection circuit designed for AdProt-based selection of leucyl-tRNA synthetase (LeuRS) activity. A premature amber stop codon (TAG) prevents translation of full-length AdProt unless LeuRS charges an amber-anticodon tRNA with leucine (pink) and suppresses the amber stop codon. (**c**) Cells were transfected with either a constitutive protease control (no selection, AdProt.FLAG), the Cre-recombinase circuit ($(LoxP)_2$Term.AdProt), or the LeuRS circuit (AdProt(STOP)) with the relevant tRNAs (pLeu-tRNA.GFP(STOP)). Transfected cells were then infected with ΔAdProt.adenovirus carrying tTA (control, tTA_{wt}.mCherry), Cre (Cre.Ad), or LeuRS (LeuRS.Ad). The infections were allowed to progress for four days before they were harvested and titered by flow cytometery. Titers are provided in infectious units per milliliter. N.D. indicates that the control virus was not detected after passaging on the synthetase selection circuit.

3.4 Discussion

We report here the development and characterization of an adaptable platform for the directed evolution of BOI functions directly in human cells. In this platform, human cells are infected by a BOI-encoding adenovirus lacking the essential AdProt and AdPol genes (**Figure 3.1c**). A newly engineered, highly error-prone variant of AdPol, EP-Pol, constitutively expressed by the human cells, replicates the adenoviral genome. The resulting error-prone DNA replication introduces mutations into the BOI gene at a high rate, thereby continuously generating mutant libraries for selection. BOI variants are then expressed during viral infection of the human cell, and continuously tested for activity via a selection couple in which functional BOI variants induce higher levels of AdProt activity stemming from an AdProt gene cassette installed in the human cells. Because AdProt activity is linked to the capacity of the virus to propagate, functional BOI variants are continuously enriched in the evolving viral population, whereas non-functional BOI variants result in non-viable virions that cannot propagate.

In developing this platform, we chose to use adenovirus rather than a natively mutagenic RNA virus owing to adenovirus' relative safety, broad tropism, ease of manipulation, and capacity to propagate even under strong selection pressure. The adenoviruses we constructed were E1-, E3-, AdPol- and AdProt-deleted. All of these genes are required for adenoviral replication in the wild. Thus, the safety of working with these adenovirus deletion variants is maximized as they can only replicate in human cells that provide essential genes in trans, and cannot replicate in unmodified human cells.[20,25,35] Moreover, the removal of this large portion of the adenoviral genome means that genes as large as ~7 kb can potentially be introduced and evolved in our platform. The broad tropism of adenovirus[36] is beneficial because it means that directed evolution experiments can, in principle, be performed in many different human cell types depending on the objective of a particular experiment.

Despite the manifold benefits of the choice to use adenovirus, we faced a significant challenge because wild-type AdPol variants has a high fidelity,[19] and therefore is unsuitable for the constant creation of mutational libraries during the course of a continuous directed evolution experiment. To address this issue, we engineered EP-Pol, a highly mutagenic AdPol variant that pushes the adenoviral mutation rate into the regime of RNA viruses such as HIV and influenza that are well-known to rapidly evolve on laboratory timescales.[24,37,38]

We used trans-complementation of EP-Pol via constitutive expression in the host cell rather than express EP-Pol directly from the adenoviral genome. By replacing the EP-Pol gene with each passage, we prevent the EP-Pol gene itself from evolving and we ensure that its mutagenic activity remains at a constant, high level throughout directed evolution experiments. We note that the optimized EP-Pol mutagenesis system may have applications beyond our directed evolution system. For instance, EP-Pol could be used to more rapidly assess resistance pathways to treatment of adenovirus infections or to improve the properties of adenovirus for therapeutic purposes.[20,39]

This mutagenesis approach does introduce mutations into the adenoviral genome outside the gene for the BOI that can potentially be negatively selected against and consequently reduce library size. The 6.5 kb genomic region we sequenced (**Figure 3.3**) was chosen because it contained both protein coding regions necessary for adenoviral replication and non-coding regions that should not face severe selection pressure.

Comparing these domains across the sequenced region, we observed only a two-fold difference between the mutation rate in the inactivated AdPol gene, which should not be under any selection pressure in our trans-complementing system, and the neighboring pIX, IVa2, and pTP genes, suggesting that such selection only impacts our mutation rate at most two-fold.

Because AdPol selectively replicates only adenoviral DNA, EP-Pol can only introduce mutations into the adenoviral genome. This mutagenesis technique thus represents an improvement over other strategies that evolve genes directly in the human genome. In such strategies, off-target mutations can arise through basal or through the enhanced mutagenesis rates, which can subvert selection pressure and generate false positives. Furthermore, even recent mutagenesis methods that target specific genes within the human genome by using somatic hypermutation[11,12] or Cas9-fusion proteins[13-15] still display significant off-target genetic modification.[40-42] Especially given the large size of the human genome, many pathways to cheating selection may be available. Our use of an orthogonal replication system means that the human host cells are discarded and replaced with each passage, preventing mutation accumulation in the human cell that could potentially cheat selection pressure. As a result, mutations that can lead to false positives are reduced to only those that occur on the ~30 kb viral genome. With a much smaller genome, we expect that the number of evolutionary escape options available to adenovirus is much fewer than those available to human cells. This advantage, combined with the much more rapid expansion of adenovirus relative to human cells, highlights the ability of our platform to quickly scan mutational space with minimal risk of selection subversion.

We found that AdProt can serve as a robust selectable marker for adenovirus-mediated directed evolution in human cells. As an enzyme with catalytic activity, we might not expect AdProt to exhibit a large dynamic range of selection. However, we observed that AdProt was able to modulate viral titers ~10-fold in response to protease levels. Importantly, we found that a small molecule inhibitor of AdProt could be easily used to further enhance this dynamic range to several orders of magnitude. It is noteworthy that the AdProt inhibitor may also be employed to actively fine-tune selection stringency over the course of a directed evolution experiment, simply by modulating the compound's concentration in cell culture media.

We used AdProt-based selection to evolve transcriptionally active variants of tTA that gained dox-resistance. Across two replicates of the experiment, two different tTA variants ultimately fixed in the population, both of which were indeed dox-resistant. We also observed a large number of lower frequency mutations at various passages above our 1% threshold for detection. The observation of these variants suggests that our platform is effectively screening sequence space for a selective advantage, particularly as the vast majority of mutations are unlikely to ever attain a frequency of 1% by chance in the evolving viral population. This proof-of-concept experiment specifically highlights how AdProt-based selection could be used to evolve transcription factors.

To explore evolving Cre recombinase, we adapted a previously reported genetic circuit[43] by placing a polyadenylation signal flanked by loxP sites upstream of the AdProt gene. The polyadenylation signal significantly attenuates AdProt expression unless it is excised by Cre recombinase. We showed that Cre-containing adenovirus infecting cells with this circuit yielded ~10-fold more progeny than a negative control virus lacking Cre,

suggesting that we could feasibly evolve tyrosine recombinases. Site specific tyrosine recombinases, such as Cre, are widely used for genetic engineering,[33] conditional knockouts in animals,[44] and constructing genetic logic gate circuits.[45] Site-specific recombinases have even been engineered to target and excise proviral HIV genomes with high specificity as a potential genetic therapy for HIV infection.[33] The ability to tailor recombinases to recognize specific sequences with high activity in human cells would both expand the recombinase genetic toolbox and allow for the facile development of new potential gene therapies.

Another interesting biotechnology application for our platform is the evolution of aminoacyl-tRNA synthetases (aaRSs) that incorporate of unnatural amino acids in human cells. Unnatural amino acid incorporation is a powerful tool for site-specifically incorporating probes and novel functionalities into proteins. However, application of this method in human cells has been limited by the dearth of tRNA/aminoacyl-tRNA synthetase (aaRS) pairs[46] and the low activity of existing aaRSs, such as pyrrolysyl-tRNA synthetases (PylRS) and others, in human cells.[47,48] Furthermore, the widely-utilized PylRSs are prone to aggregation, resulting in low intracellular concentrations of functional enzyme.[49] In order to guarantee that new evolved tRNA/aaRS pairs have optimal unnatural amino acid incorporation activity in human cells, they should be evolved directly in human cells. As a first step towards realizing this goal, we constructed a selection circuit that couples AdProt expression to the activity of Leucyl-tRNA synthetase via an AdProt gene that has a premature stop codon. We showed that cells containing this circuit only produced virus when infected with adenovirus containing the Leucyl-tRNA synthetase and did not produce any detectable virus when infected with a negative control virus. We have thus shown that evolving aaRSs with our platform should be feasible and have taken a promising step toward generating new aaRSs with high activity in human cells.

In order to further improve the platform and make it even more user-friendly, we constructed a version of a ΔAdProtΔAdPol-adenovirus genome that can be easily without the use of a complicated, advanced genetic engineering technique such as recombineering (see Chapter 2). The vector, which we termed AdEvolveDEST, contains attR1 and attR2 sites such that genes encoding a BOI can be integrated into the adenoviral genome using Gateway cloning(**Figure 3.15**).[50] Once a BOI is inserted into AdEvolveDEST, one simply generates the necessary adenovirus via plasmid transfection into the producer cell line that constitutively expresses both AdPol and AdProt. Directed evolution then simply involves serial passaging of the adenovirus on user-defined 'selector cells'.

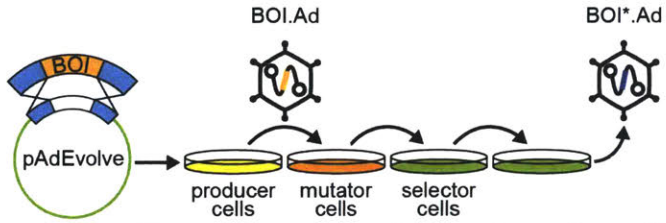

Figure 3.15 Experimental workflow for adenovirus-based directed evolution in human cells.
The gene encoding a biomolecule of interest (BOI) is first inserted into pAdEvolve. "Producer" cells (see cell lines listed in **Table 3.2**) are used to generate ΔAdProtΔAdPol-adenoviruses carrying the BOI gene. If desired, the BOI gene can be mutated prior to selection by first passaging the adenovirus on a "mutator" cell line constitutively expressing EP-Pol. A "selector" cell line tailored to the activity of interest is generated by the researcher, followed by serial passaging of viral supernatants on the selector cells.

We envision a series of future improvements that can also improve platform performance and scope. The current system relies on serial passaging of adenovirus on adherent cells. Transitioning to suspension cells would enable variant libraries several orders of magnitude larger than we can currently explore. The integration of emerging targeted mutagenesis techniques, such as MutaT7[51] or CRISPR-X,[14] could further focus mutations only to the BOI gene and also increase mutation library size. Additionally, the present system is only capable of positive selection. Implementation of a negative selection strategy would enable our platform to evolve biomolecules that have minimal residual parent activity and are more specific for a given activity.

We successfully demonstrated preliminary selection circuits to enable the directed evolution of DNA recombinases and amino-acyl tRNA synthetases, suggesting that the platform should be readily generalizable to evolve a wide variety of biological functions. Examples of the necessary selection couples already exist for an assortment of other protein classes, including TALENs,[52] proteases,[53] protein-protein interactions,[54] RNA polymerases,[17,55] Cas9,[56] and beyond. We note that phage-assisted continuous evolution in bacteria can afford larger library sizes with more tunable mutation rates, in addition to dynamic selections that occur on the order of hours, not days.[17] Critically, while adenovirus-mediated directed evolution explores mutational space more slowly than phage-assisted continuous evolution, it makes possible similar experiments in the metazoan cell environment for the first time. Thus, the platform provides a compelling option in any situation where the evolution of optimal BOI variants is unlikely to succeed in simpler systems.

3.5 Conclusion

Our platform offers several advantages relative to extant strategies for human cell-based directed evolution that rely on time-intensive screens and extensive *in vitro* manipulations. The use of adenovirus allows researchers to continuously mutate, select, and amplify genes of interest by simply transferring viral supernatant from one cell plate to the next. Owing to this simple viral passaging protocol, library sizes are restricted only by a researcher's tissue culture capacity. Cheating is minimized because mutations are specifically directed to the viral genome and any mutations that occur in the human genome are lost at the end of each passage as the adenovirus is transferred to new cells. The adenoviruses used lack multiple genes required for replication in the wild, adding a measure of safety to the system. Moreover, the user-defined nature of the selector cell and the broad tropism of adenovirus type-5 enable directed evolution to be performed in a diverse array of human cell types. We have successfully applied our adenovirus based system to the evolution of a non-viral biomolecule in human cells, and we are continuing to tune and optimize the system. By making it possible for researchers to evolve diverse BOI functions in the same environment in which the BOIs are intended to function, we believe this human cell-based directed evolution platform holds significant potential to enable us and other researchers to rapidly evolve a wide variety of biomolecules in human cells.

3.6 Materials and Methods

Cloning methods: All PCR reactions for cloning and assembling recombineering targeting cassettes were performed using Q5 High Fidelity DNA Polymerase (New England BioLabs). Restriction cloning was performed using restriction endonucleases and Quick Ligase from New England BioLabs. Adenoviral constructs were engineered using ccdB recombineering, as previously described[57] and further optimized by us.[58] Primers were obtained from Life Technologies and Sigma-Aldrich (**Table 3.4**). The TPL gene block was obtained from Integrated DNA Technologies (**Table 3.4**). Plasmid sequences can be obtained from GenBank using the accession numbers provided in **Table 3.5**. LV-Cre pLKO.1 was a gift from Elaine Fuchs (Addgene plasmid #25997)[59] and pANAP was a gift from Peter Schultz (Addgene plasmid #48696).[60]

Wild-type AdPol and EP-Pol vectors: The lentiviral vector encoding HA-tagged wild-type AdPol was previously described.[20] Mutations were introduced by site-directed mutagenesis.

CMV.AdProt vector: A 641 bp fragment containing adenoviral protease (AdProt) was amplified from the Ad5 genome using the primers BamHI.AdProt.Forward and SalI.AdProt.Reverse (**Table 3.4**) and ligated into pTRE-Tight (Clontech) using BamHI and SalI to make the pTRE-Tight.AdProt vector. The Ad5 Tripartite leader sequence (TPL) was amplified from the TPL gene block using the primers TPL.GA.Forward and TPL.GA.Reverse (**Table 3.4**) and the pTRE-Tight.AdProt vector was amplified using the primers Tight.AdProt.GA.Forward and Tight.AdProt.GA.Reverse (**Table 3.4**). The TPL and pTRE-Tight.AdProt amplicons were assembled using the HiFi DNA assembly kit (New England Biolabs) to create the pTRE-Tight.TPL.AdProt vector. From this vector, an 852 bp fragment containing TPL.AdProt was amplified using the primers NotI.TPL.AdProt.Forward and XbaI.TPL.AdProt.Reverse (**Table 3.4**) and inserted into the pENTR1A vector (Thermo Fisher) using NotI and XbaI. The LR clonase II enzyme mixture (Thermo Fisher) was used to recombine the TPL.AdProt fragment from pENTR1A.TPL.AdProt into pLenti.CMV.Hygro (w117-1) (Thermo Fisher).

pLVX.TRE3G.AdProt vector: TPL.AdProt was amplified from pTRE-Tight.TPL.AdProt using the primers TPL.AdProt.GA.Forward and TPL.AdProt.GA.Reverse (**Table 3.4**) and assembled with NotI-digested pLVX.Tight.Puro (Takara Biosciences) using the HiFi DNA assembly kit to form pLVX.Tight.TPL.AdProt.Puro. A fragment containing TPL.AdProt was obtained from pLVX.Tight.TPL.AdProt.Puro by digestion with EcoRI and BamHI and ligated into the pLVX.TRE3G vector (Takara Bio) to create the pLVX.TRE3G.AdProt vector.

pLVX.TRE3G.eGFP vector: A 762 bp fragment containing eGFP was amplified from the eGFP-N3 vector (Takara Bio) using the primers NotI.eGFP.Forward and EcoRI.eGFP.Reverse (**Table 3.4**) and ligated into the pLVX-TRE3G vector (Takara Bio) using NotI and EcoRI to create the pLVX.TRE3G.eGFP vector.

pBud.tTA.mCherry variant vectors: A 743 bp fragment containing mCherry was amplified from a pcDNA3.1-mCherry template plasmid using the primers

NotI.mCherry.Forward and XhoI.mCherry.Reverse (**Table 3.4**) and inserted into the pBudCE4.1 vector (Thermo Fisher) using NotI and XhoI to create the pBud.mCherry vector. A 771 bp fragment containing tTA was amplified from a tTA.mCherry adenoviral vector using the primers SalI.TTA.Forward and BamHI.TTA.Reverse (**Table 3.4**) and inserted into the pBud.mCherry vector using BamHI and SalI to create the pBud.tTA.mCherry vector. Site-directed mutagenesis was then performed on pBud.tTA.mCherry using a QuickChange II XL Site-Directed Mutagenesis Kit (Agilent) to generate the indicated point mutations in tTA (**Figure 3.11f**).

AdProt.FLAG vector: From the pTRE-Tight.TPL.AdProt vector, an 852 bp fragment containing TPL.AdProt was amplified using primers NotI.TPL.Forward and XbaI.AdProt.Reverse (**Table 3.4**) and inserted into the pENTR1A vector using NotI and XbaI to form pENTR1A.TPL.AdProt. A FLAG epitope tag was inserted into pENTR1A.TPL.AdProt using primers pENTR1A.AdProt.FLAG.Forward and pENTR1A.AdProt.FLAG.Reverse (**Table 3.4**) and using the QuickChange II XL Site-Directed Mutagenesis Kit to form pENTR1A.TPL.AdProt.FLAG. This vector was then recombined with pcDNA-DEST40 using LR Clonase II Enzyme Master Mix to form AdProt.FLAG.

(LoxP)$_2$Term.AdProt vector: A vector containing an SV40-polyA terminator flanked by two loxP sites was purchased from GeneArt (ThermoFisher). From this vector, a 370 bp fragment containing the floxed SV40-terminator signal was amplified using primers LoxP2Term.GA.Forward and LoxP2Term.GA.Reverse (**Table 3.4**). The pENTR1A.TPL.AdProt.FLAG vector was linearized using pENT.AdProt.GA.Forward and pENT.AdProt.GA.Reverse (**Table 3.4**). The two amplicons were assembled to form pENTR1A.(LoxP)$_2$Term.TPL.AdProt.FLAG using the NEB HiFi DNA assembly kit. This vector was then recombined with pcDNA-DEST40 using LR Clonase II Enzyme Master Mix to form (LoxP)$_2$Term.AdProt.

AdProt(STOP) vector: pENTR1A.TPL.AdProt.FLAG was mutagenized using primers L8.STOP.Forward and L8.STOP.Reverse (**Table 3.4**) to form pENTR1A.TPL.AdProt(STOP).FLAG using the QuickChange II XL Site-Directed Mutagenesis Kit. This vector was then recombined with pcDNA-DEST40 using LR Clonase II Enzyme Master Mix to form AdProt(STOP).

pLeu-tRNA.LeuRS vector: A 2607 bp fragment containing LeuRS, the *E. coli* leucyl-tRNA synthetase, was amplified from DH10B *E. coli* genomic DNA using the primers HindIII.LeuRS.Forward and XhoI.LeuRS.Reverse (**Table 3.4**) and inserted into pANAP[60] using HindIII and XhoI to create the pLeu-tRNA.LeuRS vector.

pLeu-tRNA.GFP(STOP) vector: Site-directed mutagenesis was performed on the pcDNA3.1-CMV.GFP plasmid using a QuickChange II XL Site-Directed Mutagenesis Kit (Agilent) and the primers Tyr40TAG.Forward and Tyr40TAG.Reverse (**Table 3.4**) to introduce a premature stop codon at position 40 in eGFP. Then a 750 bp fragment containing eGFP(STOP) was amplified from the site-directed mutagenesis product using the primers HindIII.eGFP.Forward and XhoI.eGFP.Reverse (**Table 3.4**) and inserted into pANAP[60] using HindIII and XhoI to create the pLeu-tRNA.GFP(STOP) vector.

Table 3.4 Primers used for cloning and recombineering.

Name	Sequence
BamHI.AdProt.Forward	5′-aaaaaaggatccaccatgggctccagtgag-3′
SalI.AdProt.Reverse	5′-aaaaagtcgacttacatgtttttcaagtgacaaaaagaag-3′
TPL.GA.Forward	5′-atcgcctggagaattcactctcttccgcatcgct-3′
TPL.GA.Reverse	5′-ctcactggagcccattgcgactgtgactggttag-3′
TPL gene block	5′-aaaaaagaattcactctcttccgcatcgctgtctgcgagggccagctgttgggctcgcggttgaggacaaactcttcgcggtctttccagtactcttggatcggaaacccgtcggcctccgaacaggtactccgccgccgagggacctgagcgagtccgcatcgaccggatcggaaaacctctcgagaaaggcgtctaaccagtcacagtcgcaggatccttttttt-3′
Tight.AdProt.GA.Forward	5′-atgggctccagtgagcag-3′
Tight.AdProt.GA.Reverse	5′-gaattctccaggcgatctg-3′
NotI.TPL.AdProt.Forward	5′-aaaaaagcggccgcactctcttccgcatcg-3′
XbaI.TPL.AdProt.Reverse	5′-aaaaaatctagattacatgtttttcaagtgacaaaaagaag-3′
TPL.AdProt.GA.Forward	5′-tggagaaggatccgcactctcttccgcatcgct-3′
TPL.AdProt.GA.Reverse	5′-atctagagccggcgcttacatgtttttcaagtgacaaaaagaag-3′
NotI.eGFP.Forward	5′-aaaaaaagcggccgccgccaccatggtgag-3′
EcoRI.eGFP.Reverse	5′-aaaaaagaattccggccgctttacttgtac-3′
NotI.mCherry.Forward	5′-aaaaaagcggccgcgcaccatggtgagcaag-3′
XhoI.mCherry.Reverse	5′-aaaaaactcgagactacttgtacagctcgtccatg-3′
SalI.TTA.Forward	5′-aaaaaagtcgacatgtctagactggacaagagcaaag-3′
BamHI.TTA.Reverse	5′-aaaaaaggatccttacccggggagcatgtcaagg-3′
NotI.TPL.Forward	5′-aaaaaagcggccgcactctcttccgcatcg-3′
XbaI.AdProt.Reverse	5′-aaaaaatctagattacatgtttttcaagtgacaaaaagaag-3′
pENTR1A.AdProt.FLAG.Forward	5′-taatctagacccagctttcttgtacaaagttggcattataag-3′
pENTR1A.AdProt.FLAG.Reverse	5′-agaaagctgggtctagattacttatcgtcgtcatccttgtaatccatgtttttcaagtgacaaaaagaagtggcg-3′
pENT.AdProt.GA.Forward	5′-gaattcgcggccgcac-3′
pENT.AdProt.GA.Reverse	5′-ggtaccggatccagtcgac-3′
LoxP2Term.GA.Forward	5′-agtcgactggatccggtaccgccgcatcaacgagctc-3′
LoxP2Term.GA.Reverse	5′-gagagtgcggccgcgaattcgaggcccagagggtacc-3′
L8.STOP.Forward	5′-cagtgagcaggaatagaaagccattgtcaaagatcttggttgtgg-3′
L8.STOP.Reverse	5′-ctttgacaatggctttctattcctgctcactggagcccattg-3′
HindIII.LeuRS.Forward	5′-aaaaaaaagcttatgcaagagcaataccgccc-3′
XhoI.LeuRS.Reverse	5′-aaaaaactcgagttagccaacgaccagattgaggag-3′
Tyr40TAG.Forward	5′-agggcgatgccacctagggcaagctg-3′
Tyr40TAG.Reverse	5′-cagcttgccctaggtggcatcgccct-3′
HindIII.eGFP.Forward	5′-aaaaaaaagcttgccaccatggtgagcaagg-3′
XhoI.eGFP.Reverse	5′-aaaaaactcgagttacttgtacagctcgtccatgcc-3′
E1.kanccdB.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccaccttaaccctcatcagtgccaacatagtaag-3′
E1.kanccdB.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaaccgctcattaggcggc-3′
E1.CMV.Promoter.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccaccttaagccacgcccacagatatacgcgttgacattg-3′
E1.bGH.polyA.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaatagaagccatagagcccac-3′
E4.kanccdB.Forward	5′-caaaaaacccacaacttcctcaaatcgtcacttccgttttcccacgttacccctcatcagtgccaacatagtaag-3′
E4.kanccdB.Reverse	5′-agtaacttgtatgtgttgggaattgtagttttcttaaaatgggaagtgacccgctcattaggcgggc-3′
E4.SV40.Promoter.Forward	5′-caaaaaacccacaacttcctcaaatcgtcacttccgttttcccacgttacttctgtggaatgtgtgtcagttaggg-3′

E4.SV40.polyA.Reverse	5′-agtaacttgtatgtgttgggaattgtagttttcttaaaatgggaagtgacctctagctagaggtcg acggtatac-3′
Pol.kanccdB.Forward	5′-tcccgcgcttcttggaactttacattgtgggccacaacatcaacggccctccctcatcagtgcca acatagtaag-3′
Pol.kanccdB.Reverse	5′-ggcacctcggaacggttgttaattacctgggcggcgagcacgatctcgtcccgctcattaggc gggc-3′
delPol.Forward	5′-gcgcggccttccggagcgaggtgtgggtgagcgcaaaggtgtccctgaccatgaccagca tgaagggcacgagctgcttcccaaaggcccccatccaag-3′
delPol.Reverse	5′-cttggatgggggcctttgggaagcagctcgtgcccttcatgctggtcatggtcagggacacctttt gcgctcacccacacctcgctccggaaggccgcgc-3′
AdProt.kanccdB.Forward	5′-ggcaacgccacaacataaagaagcaagcaacatcaacaacagctgccgccccctcatca gtgccaacatagtaag-3′
AdProt.kanccdB.Reverse	5′-tacaaataaaagcatttgcctttattgaaagtgtctctagtacattatttccgctcattaggcggg c-3′
delAdProt.Forward	5′-ggcaacgccacaacataaagaagcaagcaacatcaacaacagctgccgccaaataatgt actagagacactttcaataaaggcaaatgcttttatttgta-3′
delAdProt.Reverse	5′-tacaaataaaagcatttgcctttattgaaagtgtctctagtacattatttggcggcagctgttgttga tgttgcttgcttctttatgttgtggcgttgcc-3′
Cre.kanccdB.Forward	5′-tggaactaatcatatgtggcctggagaaacagctaaagtgcgaaagcggcccgctcattag gcgggc-3′
Cre.kanccdB.Reverse	5′-cgcgaacaaatgtggtatggctgattatgatcctctagagataattctagccctcatcagtgcca acatagtaag-3′
Cre.Forward	5′-tggctagcgtttaaacttaagcttggtacccctccgcggggatcctctaggccaccatgcccaa gaagaagaggaag-3′
Cre.Reverse	5′-cgcgaacaaatgtggtatggctgattatgatcctctagagtaattctagctaatcgccatcttcca gcagg-3′
BAC2pUC.Forward	5′-cccgggaattcggatctgc-3′
BAC2pUC.Reverse	5′-ccgggaattcggatccttgaagac-3′
AdProt.Forward	5′-gggtacccaactccatgctc-3′
AdProt.Reverse	5′-aagtggcgctcctaatctgc-3′
RPLP2.Forward	5′-ccattcagctcactgataaccttg-3′
RPLP2.Reverse	5′-cgtcgcctcctacctgct-3′

Table 3.5 Plasmid sequence accession numbers.

Vector Name	GenBank Accession Number
Wild-type AdPol vector	MH325099
EP-Pol vector	MH325100
pTRE-Tight.TPL.AdProt	MH325101
CMV.AdProt	MH325102
pLVX.TRE3G.AdProt	MH325103
R6K-kan-ccdB	MH325106
pcDNA3.1-mCherry template plasmid	MH325107
pcDNA3.1-GFP template plasmid	MH325108
pcDNA3.1-tTA template plasmid	MH325109
pcDNA3.1-tTA_{aak} template plasmid	MH325110
pcDNA3.1-KanFDEST template plasmid	MH325111
AdCFP	MH325112
tTA_{wt}.mCherry	MH325113
tTA_{mut}.GFP	MH325114
AdEvolve-DEST	MH325115
CFP.ΔAdProtΔAdPol	MH325116
pcDNA3.1-CMV.GFP	MH777595

CFP.ΔAdPol.GFP	MK044347
pLVX.TRE3G.eGFP	MH325104
pBud.tTA.mCherry	MH325105
AdProt.FLAG	MK044342
$(LoxP)_2$Term.AdProt	MK044343
AdProt(STOP)	MK044344
pLeu-tRNA.LeuRS	MH777597
pLeu-tRNA.GFP(STOP)	MH777596
Cre.Ad	MK044345
LeuRS.Ad	MK044346

Adenoviral constructs: Adenoviral constructs were engineered using *ccdB* recombineering, as previously described[57] and further modified by us,[58] in DH10B *Escherichia coli* carrying the adenovirus type-5 genome in a chloramphenicol-resistant bacterial artificial chromosome (AdBAC). Cells carrying an AdBAC were transformed with the temperature-sensitive psc101-gbaA recombineering plasmid,[57] plated on LB (Difco) agar (Alfa Aesar) with 10 μg/mL tetracycline (CalBioChem) and 10 μg/mL chloramphenicol (Alfa Aesar), and incubated for 24 h at 30 °C. Colonies were selected and grown in LB containing 10 μg/mL tetracycline and 10 μg/mL chloramphenicol overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 μg/mL tetracycline and 10 μg/mL chloramphenicol 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 L-arabinose (Chem-Impex) and L-rhamnose (Sigma Aldrich) 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 *kanR+ccdB* targeting cassette (1.8 kV, 5.8 msec, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in super optimal broth with catabolite repression (SOC; Teknova) with 2 mg/mL L-arabinose at 30 °C for 2 h, then plated on LB agar plates with 50 μg/mL kanamycin (Alfa Aesar) and 2 mg/mL L-arabinose and incubated for 24 h at 30 °C. Colonies that grew under these conditions had incorporated the *kanR+ccdB* targeting cassette and were picked in triplicate and grown in LB with 50 μg/mL kanamycin and 2 mg/mL L-arabinose at 30 °C for 18–21 h. Note that the colonies were picked in triplicate because multimers of the AdBAC formed at a high rate (~30–50% of colonies) during the first recombineering step. Such multimers cannot be successfully recombineered in the next step. Picking three colonies and recombineering them separately in parallel increases the chances of picking a monomer that can be successfully recombineered. The cultures were then diluted 25-fold in LB with 50 μg/mL kanamycin and 2 mg/mL L-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 L-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 *kanR+ccdB* cassette

currently integrated in the genome (1.8 kV, 5.8 msec, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in SOC with 2 mg/mL L-arabinose at 30 °C for 2 h, and then were washed once with LB to remove the L-arabinose and prevent continued production of the ccdA antitoxin. The cultures were then plated on LB agar plates at various dilutions with 10 μg/mL tetracycline and 10 μg/mL chloramphenicol 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, in principle, have the desired final targeting cassette integrated, but were always screened by PCR or sequencing to confirm cassette integration as some colonies may simply inactivate the ccdB toxin.

The following modifications (**Table 3.6**) were made using the primers in **Table 3.4** to obtain the adenoviruses (**Table 3.1**) used in this work:

Table 3.6 Modifications made to the adenoviral genome.

Modification	Genotype	*Kan^R+ccdB* cassette primers used with R6K-kan-ccdB template plasmid (unless stated otherwise)	Final targeting cassette oligos or primers and template (if applicable)	Purpose of modification
AdPol Deletion	ΔAdPol	Pol.kanccdB.Forward and Pol.kanccdB.Reverse	delPol.Forward and delPol.Reverse (annealed oligos)	To prevent evolution of the adenoviral polymerase. The error-prone version was expressed in trans.
Insertion of mCherry	E4R-mCherry	E4.kanccdB.Forward and E4.kanccdB.Reverse	E4.SV40.Promoter.Forward and E4.SV40.Reverse were used to amplify from pcDNA3.1-mCherry template plasmid	mCherry was inserted to enable the visualization of infected cells. The E4 position with the rightward facing orientation was previously shown to allow for optimal expression and viral titer.[61]
Insertion of eGFP	E4R-eGFP	E4.kanccdB.Forward and E4.kanccdB.Reverse	E4.SV40.Promoter.Forward and E4.SV40.Reverse were used to amplify from pcDNA3.1-eGFP template plasmid	eGFP was inserted to enable the visualization of infected cells. The E4

				position with the rightward facing orientation was previously shown to allow for optimal expression and viral titer.[61]
AdProt Deletion	ΔAdProt	AdProt.kanccdB.Forward and AdProt.kanccdB.Reverse	delAdProt.Forward and delAdProt.Reverse (annealed oligos)	This essential viral gene was deleted so that viral replication could become dependent on the conditional expression of the adenoviral protease in trans.
Insertion of active tTA	E1L-tTA	E1.kanccdB.Forward and E1.kanccdB.Reverse	E1.CMV.Promoter Forward and E1.bGH.polyA.Reverse used to amplify from pcDNA3.1-tTA template plasmid	tTA was inserted as the evolution target that must evolve to express adenoviral protease from the host genome for efficient viral propagation. The E1 position with the leftward facing orientation was previously shown to allow for optimal expression and viral titer.[61]
Insertion of DEST cassette	E1L-DEST	E1L.kanccdB.Forward and E1.kanccdB.Reverse used to amplify from pcDNA3.1-KanFDEST template plasmid	Not applicable, only the first step is required	Insertion of a DEST cassette into the E1 position with the leftward

				facing orientation. The DEST cassette has *att*R sites that allow users to insert genes via Gateway cloning.
Replacement of the low copy BAC origin with the high copy pUC origin	Not applicable	N/A, the replacement is a one step recombineering since the origin switches from chloramphenicol to ampicillin resistant	BAC2pUC.Forward and BAC2pUC.Reverse used to amplify the origin cassette from pAd/CMV/V5-DEST (Thermo Fisher).	Switching to a high copy pUC origin allowed for the preparation of concentrated, purified DNA, which was necessary for transfection and successful adenovirus production.
Insertion of Cre Recombinase	E1L-Cre	Cre.kanccdB.Forward and Cre.kanccdB.Reverse	Cre.Forward and Cre.Reverse used to amplify from LV-Cre pLKO.1 (Addgene #25997)	Cre recombinase was inserted as a model BOI for selection circuit experiments.
Insertion of LeuRS	E1L-LeuRS	E1L.kanccdB.Forward and E1.kanccdB.Reverse used to amplify from pcDNA3.1-KanFDEST template plasmid	E1.CMV.Promoter.Forward and E1.bGH.polyA.Reverse used to amplify from pLeu-tRNA.LeuRS	LeuRS aminoacyl-tRNA synthetase was inserted as a model BOI for selection circuit experiments.

Once a clone with all of the desired genetic changes was found and confirmed by Sanger sequencing, the AdBAC single-copy replication origin was replaced with the high copy pUC origin. The cells with the correct clone were grown in LB containing 10 μg/mL tetracycline and 10 μg/mL chloramphenicol overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 μg/mL tetracycline and 10 μg/mL chloramphenicol and grown at 30 °C for ~2 h until attaining an OD_{600} of 0.3–0.4. The recombineering machinery was then induced by adding L-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 pUC origin cassette (1.8 kV, 5.8 ms, 0.1 cm cuvette; BioRad Micropulser). The

cells were recovered in SOC at 30 °C for 2 h, then plated on LB agar plates with 100 µg/mL ampicillin and incubated for 24 h at 37 °C. The resulting amp-resistant colonies should have the pUC origin inserted and were checked by verifying expected restriction digestion patterns. The colonies were grown in 25 mL LB containing 100 µg/mL ampicillin and the DNA was purified using the ZymoPURE II plasmid midiprep kit (Zymo Research) according to the manufacturer's instructions. The DNA was digested with PacI overnight at 37 °C in order to liberate and linearize the adenoviral genome. The linearized DNA was purified using the E.Z.N.A. cycle pure kit (Omega Bio-tek) according to the manufacturer's instructions.

Adenovirus production: Adenoviruses were produced by transfecting a PacI (New England BioLabs)-linearized vector into appropriate trans-complementing HEK293A cells (ΔAdPol adenoviruses on wild-type AdPol cells, ΔAdProtΔAdPol adenoviruses on producer cells; see **Table 3.2**). 24 µg of PacI-linearized adenovirus vectors mixed with 144 µL polyethyleneimine (Polysciences) in 1 mL OptiMEM (Gibco) was added to a 15 cm plate of producer cells (**Table 3.2**; ~3 × 10^7 cells). Media was replaced 8 h post-transfection, and then intermittently replaced every 2–3 days until plaques were observed (typically ~3 weeks). Once plaques were detected, cytopathic effect was observed in all cells within 5 days. Upon complete cytopathic effect, the cells and media were harvested and subjected to three freeze/thaw cycles. The cell debris was removed by centrifugation at 3,200 × g for 15 min and the supernatant stored at –80 °C.

Cell culture: Cells were cultured at 37 °C and 5% CO_2(g). New cell lines were derived from a parent HEK293A cell line (ATCC) and cultured in Dulbecco's modified Eagle's medium (DMEM; Cellgro) supplemented with 10% fetal bovine serum (FBS; Cellgro), 1% penicillin-streptomycin (Cellgro), and 1% L-glutamine (Cellgro). For assays involving the tetracycline (Tet)-dependent transcriptional activation system, Tet-approved FBS (Takara Bio) was used. The producer and mutator cell lines (**Table 3.2**) were cultured in 50 µg/mL hygromycin (Thermo Fisher) to stably maintain transgenes, while the selector and phenotyping cell lines (**Table 3.2**) were cultured in 1 µg/mL puromycin (Corning) for the same purpose.

Continuous tTA evolution workflow: Before initiating directed evolution, 500 µL of a tTA_{wt}.mCherry adenovirus was amplified on mutator cells (see **Table 3.2**) to create a diverse viral population. After 5 days, cytopathic effect was observed in all cells. This amplified virus was harvested with three freeze/thaw cycles. Three 15 cm, semi-confluent dishes of selector cells (**Table 3.2**) (~1 × 10^7 cells/plate) were infected with either 250, 500, or 1,000 µL of the amplified virus in the presence of dox. Plates were monitored for plaques every day. If more than one plate displayed a plaque on the same day, the plate with the lowest volume of virus added was used for the next round of evolution. The day after a plaque was observed, typically every 4–8 days, three 15 cm semi-confluent dishes of selector cells were again infected in the presence of dox. The three dishes were infected with 250, 500, or 1,000 µL of media from the previous round by direct transfer without a freeze/thaw step. 2 mL of media were saved in Eppendorf tubes and stored at –80 °C for future analysis. In Trial 2, an additional media harvest was performed after full cytopathic effect was observed. In Trial 1, the concentration of dox was increased to 200

nM at passage 7 and then to 20 μM in passages 8–12. In Trial 2, the concentration of dox was held constant at 200 nM for all seven passages.

Measuring promoter activity of viral populations: To follow changes in promoter activity developing during Trial 1, phenotyping cells (**Table 3.2**) were plated in a 96-well plate at ~40,000 cells/well. The next day, 30 μL of media from passages 1–12 was used to infect two rows of the 96-well plate. Media was removed 5 h post-infection and replaced with media containing 0 μM or 20 μM dox. The cells were then analyzed by flow cytometry (see above for sample preparation) for simultaneous expression of mCherry, indicating that the cell was infected, and GFP, indicating that the promoter was activated by the tTA protein.

Viral genome isolation for next-generation sequencing: Using a viral DNA isolation kit (NucleoSpin Virus; Macherey-Nagel), DNA was harvested from 200 μL of the media that was saved after each round of evolution. A 1.75 kb region of DNA encompassing the CMV promoter and the tTA gene was PCR-amplified from 1 μL of the harvested DNA for 20 rounds of amplification using 5′-ctacataagacccccaccttatatattctttcc-3′ and 5′-agcgggaaaactgaataagaggaagtgaaatc-3′ forward and reverse primers, respectively. The resulting PCR product was purified and prepared for Illumina sequencing via the Nextera DNA Library Prep protocol (Illumina). 250 bp paired-end sequencing was run on a MiSeq (Illumina). Sequencing reads were aligned to the amplicon sequence, which was derived from the tTA_{wt}.mCherry adenovirus sequence using bwa mem 0.7.12-r1039 [RRID:SCR_010910]. Allele pileups were generated using samtools v1.5 mpileup [RRID:SCR_002105] with flags -d 10000000 --excl-flags 2052, and allele counts/frequencies were extracted.[62,63] Each position within the tTA gene and CMV promoter had at least 1,000-fold coverage.

Reverse genetics of tTA variants: HEK293A cells were seeded in a 12-well plate at ~4 × 10^5 cells/well. The next day, 0.2 μg of the pBud.tTA.mCherry vector was co-transfected with 1 μg of the pLVX-TRE3G.eGFP vector using 7.2 μL of polyethyleneimine (Polysciences) and 100 μL OPTI-MEM. 8 h post-transfection, media was exchanged and 20 μM dox was added. 48 h post-transfection, cells were analyzed by flow cytometry (see above for sample preparation). Promoter activity was calculated based on the mean fluorescence intensity of GFP fluorescence, backgated for only mCherry-expressing cells. Testing of recombinase and synthetase selection circuits: HEK293A cells expressing wt-AdPol were plated at 3.5 × 10^5 cells/well in a 12-well plate. The next day, 1 μg of the plasmid for each circuit ($(LoxP)_2$Term.AdProt, AdProt(STOP), or AdProt.FLAG as a positive control) was transfected into six wells of a 12-well plate using 6 μL of polyethyleneimine in 100 μL of OPTI-MEM. For the AdProt(STOP) circuit, 0.5 μg was co-transfected with 0.5 μg pLeu-tRNA.GFP(STOP). Media was changed 4 h post-transfection. The next day, transfected wells were infected with either the relevant BOI virus (**Table 3.1**; Cre.Ad for $(LoxP)_2$Term.AdProt, and LeuRS.Ad for AdProt(STOP)) or TTA_{wt}.mCherry as a negative control at an MOI of 5. Cells were washed 3× with media 3 h post-infection. After 4 days, media and cells were harvested and subject to three freeze/thaw cycles, followed by analysis of titers using flow cytometry.

Generation of cell lines by lentiviral transduction: In a typical protocol, ~9 × 10^6 HEK293FT cells (Thermo Fisher) were plated on a poly-D-lysine-coated 10 cm plate. The next day, the cells were co-transfected with plasmids from a third-generation lentiviral packaging system:[64] 15 µg RRE, 6 µg REV, 3 µg VSVG, and 15 µg transfer vector using 60 µL Lipofectamine 2000 (Thermo Fisher). Cultures were maintained in 5 mL total volume of OPTI-MEM (Gibco) during the transfection. After 8 h, the media was exchanged for fresh DMEM. After 48 h, media was harvested and centrifuged for 5 min at 3,200 × g to clear the cell debris. The supernatant was used to transduce HEK293A cells supplemented with 4 µg/mL polybrene (Sigma-Aldrich). After 24 h, the media was exchanged for fresh DMEM. After 48 h, media was exchanged again for DMEM containing appropriate antibiotics to select stable cell lines.

Mutagenesis rate determination: The mutagenic potential of AdPol variants was evaluated following a previously reported protocol.[20] Briefly, a polymerase-deleted Ad5, AdGLΔPol, was subjected to 10 serial passages on cultures of 911 cells[65] expressing EP-Pol in order to accumulate mutations. After 10 serial passages, 911 cells expressing wild-type AdPol were infected in duplicate 6-well plates at ~50 plaque-forming units/well in order to amplify pools of 50 viral clones for sequencing. Based on a plaque assay of one of the duplicates (which was overlaid with agarose), the actual number of plaque-forming viral clones in the pool obtained from the other duplicate (which was not overlaid with agarose) was estimated to be ~27. Using pools of 50 or fewer clonal viruses ensured that mutations present in only one clone will be present at a frequency above the threshold of detection. From the ~27-clone viral pool, a 6.5-kb fragment was amplified and prepared for deep sequencing. Libraries were subjected to 32 cycles of single-read sequencing by an Illumina Genome Analyzer II. Using the short read analysis pipeline SHORE,[66] these reads were mapped against the reference sequence allowing up to two mismatches or gaps, after which low quality base calls within the obtained mappings were individually masked. Mutations were subsequently scored using a minimal variant frequency requirement of 0.25% and a minimal local sequencing depth requirement of 1200 for both the forward and the reverse read mappings. Previous experiments showed that these settings were able to account for sequencing errors and accurately score mutations.[20]

AdPol and AdProt trans-complementation assays: The day before beginning the assay, a 6-well plate was seeded with ~1 × 10^6 of the indicated cells. The next day, individual wells were infected with the indicated adenoviruses at a low MOI (< 0.5) in order to permit observation of the presence or absence of a spreading infection. AdPol and EP-Pol trans-complementation (see **Figure 3.4** for AdPol and **Figure 3.3b** for EP-Pol) was tested by monitoring CFP.ΔAdPol.GFP adenovirus infection on either AdPol- or EP-Pol-expressing HEK293A cells. Pictures were taken with an Olympus U-TB190 microscope. AdProt and AdPol double trans-complementation (see **Figure 3.5**) was tested by monitoring CFP.ΔAdProtΔAdPol (**Table 3.1**) infection on producer cells. Pictures were taken with a Nikon Eclipse TE200 microscope.

Determining adenoviral titer by flow cytometry: Adenoviral titers were determined through flow cytometry. Known volumes of AdPol- and AdProt-deleted viral supernatants were added to AdPol-expressing HEK293A cells. 2–3 days post-infection, cells were

washed once with media, stained with 0.2 µg/mL DAPI, and then analyzed on a BD LSR II Analyzer for fluorescent protein expression. Infectious titers were determined by measuring the percentage of cells infected by a known volume of virus. To minimize counting cells that were infected by more than one virus and to minimize any background fluorescence, data were only considered if they fell within the linear range, which typically encompassed samples where 1–10% of cells were infected.

Competition experiments: A confluent dish of selector cells (**Table 3.2**; ~15 million cells) was infected with either a 1:100 or 1:1,000 mixture of tTA_{wt}:tTA_{mut} adenovirus (MOI ~ 0.25; **Table 3.1**). Plates were monitored for the appearance of spreading infection, defined by fluorescent "comets" or plaques, every 24 h. One day after the observation of spreading infection, 1 mL of media was transferred to a new semi-confluent dish (~1 × 10^7 cells) of selector cells for the next passage (see **Table 3.2**), and 2 mL of media was stored at –80 °C for later analysis. To analyze the relative amounts of each virus present after each passage, we measured the relative adenoviral titers by flow cytometry (see above). The ratio of tTA_{wt} and tTA_{mut} viruses was determined by taking the ratio of cells expressing only mCherry and only GFP.

AdProt inhibitor experiments: A confluent 12-well plate of selector cells (**Table 3.2**) (~4 × 10^5 cells/well) was infected with tTA_{wt}.mCherry adenovirus (MOI ~ 5). After 4 h, the cells were washed with PBS (Corning), and the AdProt inhibitor was added at the indicated concentrations (0 µM, 1 µM, 20 µM) in the absence or presence of 2 nM doxycycline (dox; Sigma-Aldrich). After 6 days, media and cells were harvested and subjected to three freeze/thaw cycles, and analyzed by flow cytometry (see above).
AdProt inhibitor toxicity assay: A 96-well plate of HEK293A cells were treated with the AdProt inhibitor at concentrations up to 20 µM for 5 days (**Figure 3.10**). A CellTiter-Glo Luminescent Cell Viability Assay (Promega) was performed according to the manufacturer's instructions. Readings were normalized to the 0 µM AdProt inhibitor samples.

RT-qPCR on selector cells: A confluent 12-well plate of selector cells (**Table 3.2**; ~4 × 10^5 cells/well) was transfected with 1.25 µg of pTet-Off Advanced (Takara Bio) using 7.5 µL of polyethyleneimine (Polysciences) and 100 µL OPTI-MEM. 2 days later, cells were harvested and the RNA was extracted using an E.Z.N.A Total RNA Kit (Omega Bio-Tek). cDNA was prepared from 1 µg of purified RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qPCR analysis for AdProt (primers: AdProt.Forward and AdProt.Reverse) and the housekeeping gene RPLP2 (primers: RPLP2.Forward and RPLP2.Reverse;**Table 3.4**) on a LightCycler 480 II (Roche). AdProt transcript levels were normalized to untransfected selector cells (**Table 3.2**).

Dox dose-response experiment: A confluent 24-well plate of selector cells (**Table 3.2**; ~1.5 × 10^5 cells/well) was infected with tTA_{wt}.mCherry adenovirus (MOI ~5). After 4 h, the cells were washed with DMEM (Corning), and dox was added at the indicated concentrations (0 nM, 0.02 nM, 0.1 nM, 0.2 nM, 1 nM, or 2 nM). After 5 days, media and cells were harvested and subjected to three freeze/thaw cycles, followed by analysis of titers using flow cytometry.

3.7 Acknowledgements

This work was supported by an NIH Director's New Innovator Award (1DP2GM119162), MISTI Global Seeds Fund, and MIT (all to M.D.S.) and by the European Union through the 6th Framework Program GIANT (contract no. 512087; to R.C.H.). L.J.P., C.L.M., and S.J.H. were supported by National Science Foundation Graduate Research Fellowships under Grant No. 1122374. C.M.B. was supported by an NIGMS/NIH Inter-Departmental Biotechnology Training Program (T32-GM008334). This work was also supported in part by the NIH/NIEHS under award P30-ES002109 and by Cancer Center Support (core) Grant P30-CA14051 from the NIH/NCI.

3.8 References

[1] Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins, *Nat. Rev. Genet. 16*, 379–394.

[2] Gai, S. A., and Wittrup, K. D. (2007) Yeast surface display for protein engineering and characterization, *Curr. Opin. Struct. Biol. 17*, 467–473.

[3] Romero, P. A., and Arnold, F. H. (2009) Exploring protein fitness landscapes by directed evolution, *Nat. Rev. Mol. Cell Biol. 10*, 866–876.

[4] Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein, *Nat. Biotechnol. 22*, 1567–1572.

[5] Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N., and Ting, A. Y. (2018) Efficient proximity labeling in living cells and organisms with TurboID, *Nat. Biotechnol.*, 880–887.

[6] Arzumanyan, G. A., Gabriel, K. N., Ravikumar, A., Javanpour, A. A., and Liu, C. C. (2018) Mutually orthogonal DNA replication systems *in vivo*, *ACS Synth. Biol. 7*, 1722–1729.

[7] Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system, *Cell 163*, 759–771.

[8] Peck, S. H., Chen, I., and Liu, D. R. (2011) Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells, *Chem. Biol. 18*, 619–630.

[9] Piatkevich, K. D., Jung, E. E., Straub, C., Linghu, C., Park, D., Suk, H. J., Hochbaum, D. R., Goodwin, D., Pnevmatikakis, E., Pak, N., Kawashima, T., Yang, C. T., Rhoades, J. L., Shemesh, O., Asano, S., Yoon, Y. G., Freifeld, L., Saulnier, J. L., Riegler, C., Engert, F., Hughes, T., Drobizhev, M., Szabo, B., Ahrens, M. B., Flavell, S. W., Sabatini, B. L., and Boyden, E. S. (2018) A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters, *Nat. Chem. Biol. 14*, 352–360.

[10] Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G., and Wandless, T. J. (2006) A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules, *Cell 126*, 995–1004.

[11] Wang, C. L., Yang, D. C., and Wabl, M. (2004) Directed molecular evolution by somatic hypermutation, *Protein Eng. Des. Sel. 17*, 659–664.

[12] Wang, L., Jackson, W. C., Steinbach, P. A., and Tsien, R. Y. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation, *Proc. Natl. Acad. Sci. U.S.A. 101*, 16745–16749.

[13] Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells, *Nat. Methods 13*, 1029–1035.

[14] Hess, G. T., Fresard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., Montgomery, S. B., and Bassik, M. C. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells, *Nat. Methods 13*, 1036–1042.

[15] Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, *Nature 533*, 420–424.
[16] Das, A. T., Zhou, X., Vink, M., Klaver, B., Verhoef, K., Marzio, G., and Berkhout, B. (2004) Viral evolution as a tool to improve the tetracycline-regulated gene expression system, *J. Biol. Chem. 279*, 18776–18782.
[17] Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules, *Nature 472*, 499–503.
[18] Hoeben, R. C., and Uil, T. G. (2013) Adenovirus DNA replication, *Cold Spring Harb Perspect Biol 5*, a013003.
[19] Risso-Ballester, J., Cuevas, J. M., and Sanjuán, R. (2016) Genome-wide estimation of the spontaneous mutation rate of human adenovirus 5 by high-fidelity deep sequencing, *PLOS Pathog. 12*, e1006013.
[20] Uil, T. G., Vellinga, J., de Vrij, J., van den Hengel, S. K., Rabelink, M. J., Cramer, S. J., Eekels, J. J., Ariyurek, Y., van Galen, M., and Hoeben, R. C. (2011) Directed adenovirus evolution using engineered mutator viral polymerases, *Nucleic Acids Res. 39*, e30.
[21] Kamtekar, S., Berman, A. J., Wang, J., Lazaro, J. M., de Vega, M., Blanco, L., Salas, M., and Steitz, T. A. (2004) Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29, *Mol. Cell 16*, 609–618.
[22] Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., and Belshaw, R. (2010) Viral mutation rates, *J. Virol. 84*, 9733–9748.
[23] Davis, J. N., and van den Pol, A. N. (2010) Viral mutagenesis as a means for generating novel proteins, *J. Virol. 84*, 1625–1630.
[24] Phillips, A. M., Gonzalez, L. O., Nekongo, E. E., Ponomarenko, A. I., McHugh, S. M., Butty, V. L., Levine, S. S., Lin, Y. S., Mirny, L. A., and Shoulders, M. D. (2017) Host proteostasis modulates influenza evolution, *eLife 6*, e28652.
[25] Elahi, S. M., Oualikene, W., Naghdi, L., O'Connor-McCourt, M., and Massie, B. (2002) Adenovirus-based libraries: efficient generation of recombinant adenoviruses by positive selection with the adenovirus protease, *Gene Ther. 9*, 1238-1246.
[26] Greber, U. F., Webster, P., Weber, J., and Helenius, A. (1996) The role of the adenovirus protease in virus entry into cells, *EMBO J. 15*, 1766–1777.
[27] Webster, A., Leith, I. R., and Hay, R. T. (1994) Activation of adenovirus-coded protease and processing of preterminal protein, *J. Virol. 68*, 7292–7300.
[28] Gossen, M., and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline responsive promoters, *Proc. Natl. Acad. Sci. U.S.A. 89*, 5547–5551.
[29] Loew, R., Heinz, N., Hampf, M., Bujard, H., and Gossen, M. (2010) Improved Tet-responsive promoters with minimized background expression, *BMC Biotechnol. 10*, 81.
[30] Krueger, M., Scholz, O., Wisshak, S., and Hillen, W. (2007) Engineered Tet repressors with recognition specificity for the tetO-4C5G operator variant, *Gene 404*, 93–100.
[31] Grosche, P., Sirockin, F., Mac Sweeney, A., Ramage, P., Erbel, P., Melkko, S., Bernardi, A., Hughes, N., Ellis, D., Combrink, K. D., Jarousse, N., and Altmann, E.

(2015) Structure-based design and optimization of potent inhibitors of the adenoviral protease, *Bioorg. Med. Chem. Lett. 25*, 438–443.
[32] Hecht, B., Muller, G., and Hillen, W. (1993) Noninducible Tet repressor mutations map from the operator binding motif to the C terminus, *J. Bacteriol. 175*, 1206–1210.
[33] Meinke, G., Bohm, A., Hauber, J., Pisabarro, M. T., and Buchholz, F. (2016) Cre recombinase and other tyrosine recombinases, *Chem. Rev. 116*, 12785–12820.
[34] Italia, J. S., Zheng, Y., Kelemen, R. E., Erickson, S. B., Addy, P. S., and Chatterjee, A. (2017) Expanding the genetic code of mammalian cells, *Biochem. Soc. Trans. 45*, 555–562.
[35] Russell, W. C. (2000) Update on adenovirus and its vectors, *J. Gen. Virol. 81*, 2573–2604.
[36] Lucher, L. A. (1995) Abortive adenovirus infection and host range determinants, *Curr. Top. Microbiol. Immunol. 199 (Pt 1)*, 119–152.
[37] Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Månson, L., Asjö, B., and Wain-Hobson, S. (1989) Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations, *Cell 58*, 901–910.
[38] O'Loughlin, T. L., Greene, D. N., and Matsumura, I. (2006) Diversification and specialization of HIV protease function during in vitro evolution, *Mol. Biol. Evol. 23*, 764–772.
[39] Myers, N. D., Skorohodova, K. V., Gounder, A. P., and Smith, J. G. (2013) Directed evolution of mutator adenoviruses resistant to antibody neutralization, *J. Virol. 87*, 6047–6050.
[40] Meng, F. L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K. R., Meyers, R. M., Amor, C., Wasserman, C. R., Neuberg, D., Casellas, R., Nussenzweig, M. C., Bradner, J. E., Liu, X. S., and Alt, F. W. (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability, *Cell 159*, 1538–1548.
[41] Kim, D., Lim, K., Kim, S. T., Yoon, S. H., Kim, K., Ryu, S. M., and Kim, J. S. (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases, *Nat. Biotechnol. 35*, 475–480.
[42] Wang, C. L., Harper, R. A., and Wabl, M. (2004) Genome-wide somatic hypermutation, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7352–7356.
[43] Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H., and Liu, D. R. (2016) A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells, *Nucleic Acids Res. 44*, 9758–9770.
[44] Karimova, M., Baker, O., Camgoz, A., Naumann, R., Buchholz, F., and Anastassiadis, K. (2018) A single reporter mouse line for Vika, Flp, Dre, and Cre-recombination, *Sci. Rep. 8*, 14453.
[45] Weinberg, B. H., Pham, N. T. H., Caraballo, L. D., Lozanoski, T., Engel, A., Bhatia, S., and Wong, W. W. (2017) Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells, *Nat. Biotechnol. 35*, 453–462.
[46] Zheng, Y., Mukherjee, R., Chin, M. A., Igo, P., Gilgenast, M. J., and Chatterjee, A. (2018) Expanding the scope of single- and double-noncanonical amino acid mutagenesis in mammalian cells using orthogonal polyspecific leucyl-tRNA synthetases, *Biochemistry 57*, 441–445.

[47] Yamaguchi, A., Iraha, F., Ohtake, K., and Sakamoto, K. (2018) Pyrrolysyl-tRNA synthetase with a unique architecture enhances the availability of lysine derivatives in synthetic genetic codes, *Molecules 23*, 2460.
[48] Wang, W., Takimoto, J. K., Louie, G. V., Baiga, T. J., Noel, J. P., Lee, K. F., Slesinger, P. A., and Wang, L. (2007) Genetically encoding unnatural amino acids for cellular and neuronal studies, *Nat. Neurosci. 10*, 1063–1072.
[49] Sharma, V., Zeng, Y., Wang, W. W., Qiao, Y., Kurra, Y., and Liu, W. R. (2018) Evolving the N-terminal domain of pyrrolysyl-tRNA synthetase for improved incorporation of noncanonical amino acids, *ChemBioChem 19*, 26–30.
[50] Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000) DNA cloning using *in vitro* site-specific recombination, *Genome Res. 10*, 1788–1795.
[51] Moore, C. L., Papa, L. J., III, and Shoulders, M. D. (2018) A processive protein chimera introduces mutations across defined DNA regions *in vivo*, *J. Am. Chem. Soc. 140*, 11560–11564.
[52] Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., Tsai, S. Q., Sander, J. D., Joung, J. K., and Liu, D. R. (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity, *Nat. Methods. 12*, 939–942.
[53] Dickinson, B. C., Packer, M. S., Badran, A. H., and Liu, D. R. (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations, *Nat. Commun. 5*, 5352.
[54] Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., Wang, P., Malvar, T., and Liu, D. R. (2016) Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance, *Nature 533*, 58–63.
[55] Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. (2014) Negative selection and stringency modulation in phage-assisted continuous evolution, *Nat. Chem. Biol. 10*, 216–222.
[56] Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, *Nature 556*, 57–63.
[57] Wang, H., Bian, X., Xia, L., Ding, X., Muller, R., Zhang, Y., Fu, J., and Stewart, A. F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection, *Nucleic Acids Res. 42*, e37.
[58] Papa, L. J., III, and Shoulders, M. D. (2019) Genetic Engineering by DNA Recombineering, *Curr. Protoc. Chem. Biol. 11*, e70.
[59] Beronja, S., Livshits, G., Williams, S., and Fuchs, E. (2010) Rapid functional dissection of genetic networks via tissue-specific transduction and RNAi in mouse embryos, *Nat. Med. 16*, 821–827.
[60] Chatterjee, A., Guo, J., Lee, H. S., and Schultz, P. G. (2013) A genetically encoded fluorescent probe in mammalian cells, *J. Am. Chem. Soc. 135*, 12540–12543.
[61] Suzuki, M., Kondo, S., Pei, Z., Maekawa, A., Saito, I., and Kanegae, Y. (2015) Preferable sites and orientations of transgene inserted in the adenovirus vector genome: The E3 site may be unfavorable for transgene position, *Gene Ther. 22*, 421–429.

[62] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools, *Bioinformatics 25*, 2078–2079.

[63] Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics 27*, 2987–2993.

[64] Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., and Naldini, L. (1998) A third-generation lentivirus vector with a conditional packaging system, *J. Virol. 72*, 8463–8471.

[65] Fallaux, F. J., Kranenburg, O., Cramer, S. J., Houweling, A., Van Ormondt, H., Hoeben, R. C., and Van Der Eb, A. J. (1996) Characterization of 911: a new helper cell line for the titration and propagation of early region 1-deleted adenoviral vectors, *Hum. Gene Ther. 7*, 215–222.

[66] Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N., and Weigel, D. (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads, *Genome Res. 18*, 2024–2033.

Chapter 4: Targeted Mutagenesis by Fusing T7 RNA Polymerase to DNA Damaging Enzymes

Portions of the work presented in this chapter have been adapted from the following manuscript and appear in the thesis of Dr. Christopher Lawrence Moore:

Moore, C. L.*; Papa, L. J., III*; Shoulders, M. D. "A processive protein chimera introduces mutations across defined DNA regions *in vivo*" *Journal of the American Chemical Society* **2018**, 140, 11560–11564.

*Denotes equal contributions (authors were 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. L.J.P. constructed and characterized the tadA variants, with assistance from Amanuella A. Mengiste. The MIT BioMicro Center performed Illumina sequencing. We thank Vincent L. Butty for downstream analysis of sequencing data. M.D.S. supervised research and acquired funding.

4.2 Introduction

Traditional *in vivo* mutagenesis strategies, which are critical for studying and using evolution in living systems, rely on exogenous mutagens (e.g., high energy light or chemicals)[1,2] or expressing mutagenic enzymes (e.g., XL1-Red[3] or the MP6 plasmid[4]). These global mutagenesis strategies can yield high mutation rates and diverse genetic landscapes. However, extensive mutations throughout the genome are problematic in many contexts, especially in directed evolution experiments (**Figure 4.1**). Off-target mutations outside the intended DNA region are often toxic when they occur in essential portions of the genome,[5,6] a problem that limits library size and engenders rapid silencing of mutagenic plasmids. Global mutagens also introduce "parasite" variants into DNA libraries, originating from mutations outside the gene of interest that allow an organism to circumvent selection schemes.[7]

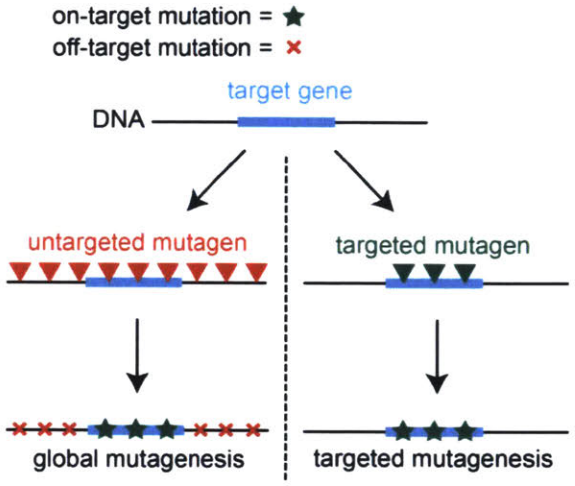

Figure 4.1 Global versus targeted mutagenesis.
In global mutagenesis, only a small fraction of mutations introduced occur in the target gene.

Targeted *in vivo* mutagenesis strategies have the potential to overcome these deficiencies. DNA-damaging enzymes fused to deactivated Cas9 nucleases can edit bases at specific genetic loci,[8-12] but require many gRNAs to tile mutagenic enzymes throughout a target DNA that may be multi-kb in length.[13,14] Moreover, the guide RNAs must be redesigned after each evolution round introduces new mutations in the target DNA. Another example is the use of an error-prone polI variant to selectively mutagenize genes on ColE1 plasmids, although this method is limited to *Escherichia coli* and can target mutations within only a few kb of the ColE1 origin.[15,16] Error-prone replication mediated by the Ty1 retrotransposon specifically in yeast can also selectively mutate <5 kb genetic cargoes inserted into the retrotransposon.[17] Other targeted mutation methods in yeast include oligo-mediated genome engineering,[18] which can be labor-intensive, and an orthogonal replication system,[19] which was developed specifically in yeast.

We rationalized that a processive, DNA-traversing biomolecule tethered to a DNA-damaging enzyme could provide a generalizable solution to the problem of targeting mutations across large, yet still well-defined, DNA regions. Monomeric RNA polymerases possess inherently high promoter specificity[20] and processivity.[21] Cytidine deaminases are potent DNA-damaging enzymes that can act on single-stranded DNA substrates

during transcription.[22] We envisioned that a chimeric "MutaT7" protein consisting of a cytidine deaminase (rApo1) fused to T7 RNA polymerase would, therefore, allow us to target mutations specifically to any DNA region lying downstream of a T7 promoter (**Figure 4.2**), provided the T7 promoter is not present elsewhere in the genome.

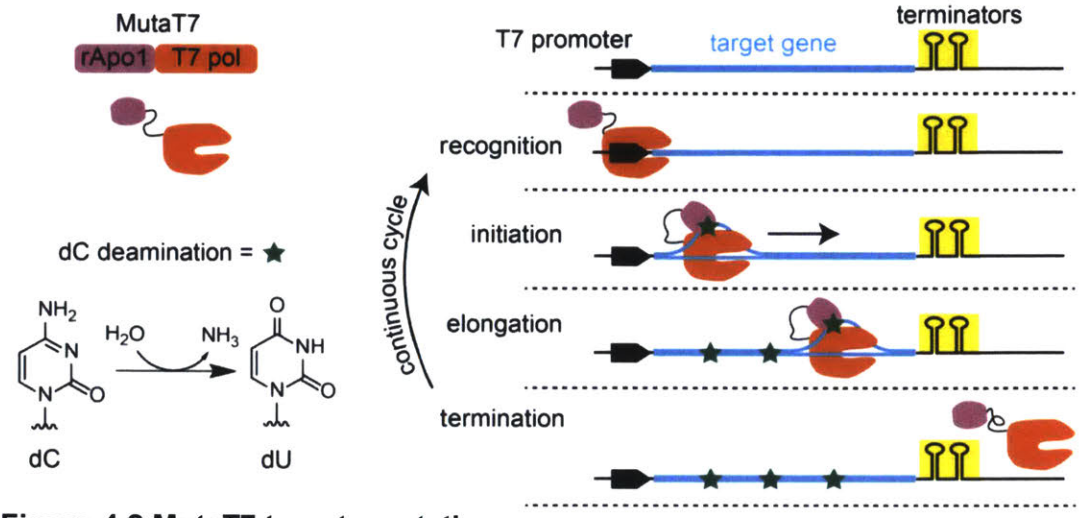

Figure 4.2 MutaT7 targets mutations.
MutaT7 utilizes a cytidine deaminase to target mutations with high specificity between a T7 promoter and a terminator array.

We demonstrate that the MutaT7 chimera is a potent and highly targeted *in vivo* mutagenesis agent. Unlike global mutagenesis methods, MutaT7 did not result in a high rate of off-target mutagenesis in the *E. coli* genome and did not significantly impact cell viability, allowing for the creation of large libraries and for the accumulation of mutations over long time scales. We further showed that the mutational spectrum can be expanded by fusing other DNA damaging enzymes, such as the adenosine deaminase tadA, to the T7 RNA polymerase. MutaT7 thus provides a mechanism for *in vivo* targeted mutagenesis across multi-kb DNA sequences. MutaT7 should prove useful in diverse organisms, opening the door to new types of *in vivo* evolution experiments.

4.3 Results

4.3.1 MutaT7 construction and validation of transcription activity

To begin, we created MutaT7 by fusing rApol to the T7 RNA polymerase using a flexible Gly-Ser linker.[23] Expressing MutaT7 from a multi-copy p15 plasmid with a strong arabinose-inducible promoter appeared to be toxic to the cells, presumably because of excessive cytidine deaminase activity. In order to lower basal expression of MutaT7, we integrated the MutaT7 gene into the genome of *E. coli* under the control of a promoter designed to display tightly-controlled, weak expression. The promoter was derived from the $P_{A1lacO-1}$ promoter, which was engineered to be tightly repressed by the lacI repressor.[24] We further weakened the strength of this promoter by replacing the σ^{70} binding sites with those of the BBa_J23114 promoter from the Anderson Collection (**Figure 4.3a**), which has about 1/10 of the strength of σ^{70} consensus binding sites.[10] Additionally, we increased the strength of the promoter of the *lacI* gene (**Figure 4.3b**), since higher *lacI* expression can result in better repression.[25] Along with minimizing toxicity by lowering the expression of MutaT7, we sought to maximize the mutagenicity of the strain (**Table 4.3**) by deleting uracil DNA glycosylase (Δ*ung*). Deleting *ung* inhibits repair of deoxyuridine to deoxycytidine and increases mutagenesis rates,[26,27] especially in the context of cytidine deaminases.[28] We also constructed an identical negative control strain (drApo1–T7, **Table 4.3**), in which the rApol domain contained an E63Q mutation known to render rApol catalytically dead.[29]

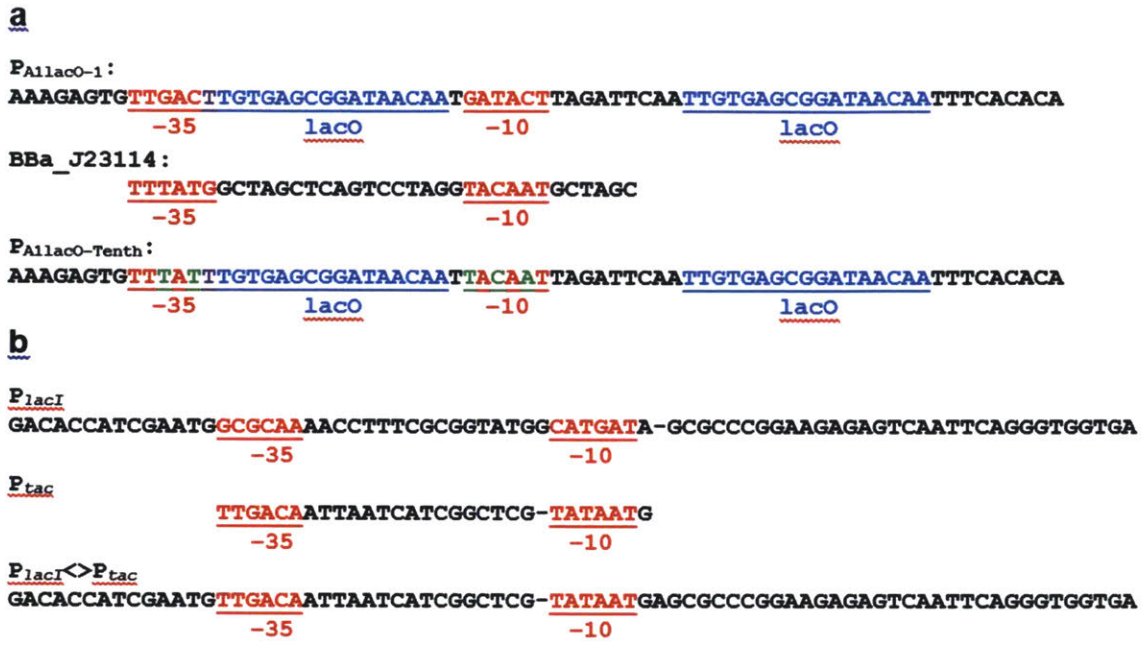

Figure 4.3 Promoter design for MutaT7 and lacI expression.
(**a**) The lac operators are shown in blue and the σ^{70} binding sites are shown in red. The BBa_J23114 promoter has about 1/10 of the strength of the σ^{70} consensus binding sites. The σ^{70} binding sites of BBa_J23114 were grafted onto $P_{A1lacO-1}$ 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.

With the drApol–T7 and MutaT7 strains constructed, we sought to test if the T7 RNA polymerase was still functional and still able to efficiently transcribe tens of kilobases[21] when fused to rApol. We utilized a lacZ expression assay[30] and a set of reporter plasmids with either no promoter or a T7 promoter near the *lacZα* gene, or 21 kb upstream the *lacZα* gene (**Figure 4.4a**). Reporters with the strong, constitutive P_{tac} promoter near the *lacZα* gene served as positive controls. We found that the drApol–T7 fusion was just as active as the unfused T7 RNA polymerase in the drApol+T7 *ung*$^+$ strain, and was still able to efficiently transcribe the *lacZα* from a T7 promoter 21 kb away (**Figure 4.4b**).

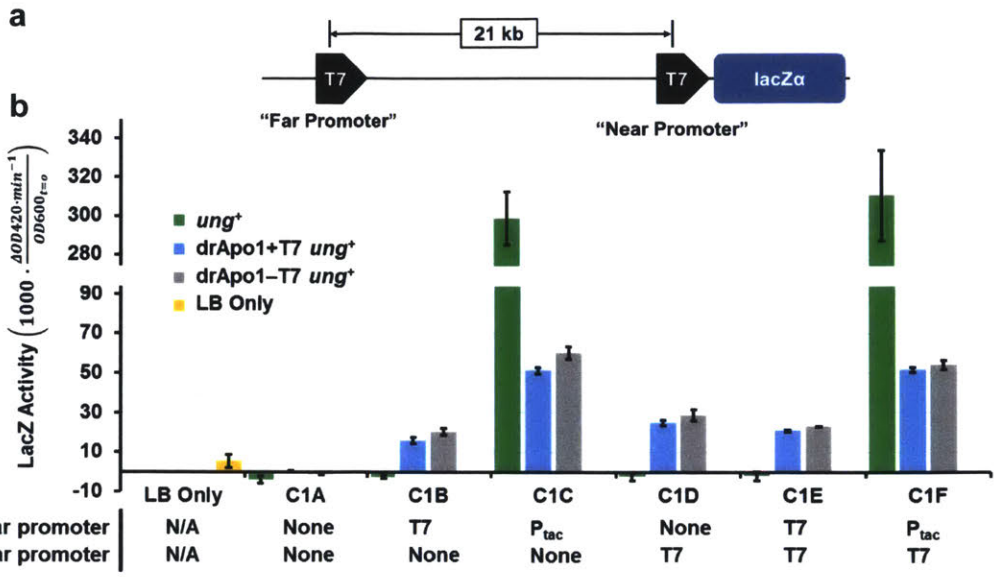

Figure 4.4 MutaT7 maintains a high level of activity and processivity.
(**a**) A general diagram of the lacZα reporter plasmid C1E, which has both a "near" and "far" T7 promoter. The rest of the lacZα reporter plasmids are missing one or both of these T7 promoters, or have a strong, constitutive P_{tac} promoter in place of the "near" T7 promoter. The genome of human adenovirus type-5 served as the intervening DNA between the "near" and "far" promoters. (**b**) LacZ activity measured via *o*NPG cleavage. *ung*$^+$ served as a negative control that lacks the T7 RNA polymerase. "drApo1+T7 *ung*$^+$" served as a positive control in which deactivated rApo1 and active T7 are expressed as separate proteins. Various reporters were used with different locations of targeted (T7) promoters and constitutive (P_{tac}) promoters, as indicated by the key on the x-axis. "LB Only" was a negative control in which LB with no cells was added to the assay mixture. Note, the *ung*$^+$ strain likely has extremely high lacZ activity because it has the endogenous P_{lacI} promoter (**Table 4.3**) and likely expresses lower levels of the lacI repressor, which represses both the $P_{A1lacO\text{-}Tenth}$ promoter and the P_{lac} promoter. Less repression of P_{lac} would mean higher expression of the lacZω fragment, which complements with the lacZα fragment in the assay to cleave *o*NPG.

4.3.2 MutaT7 targets mutations with minimal off-target activity and toxicity

Once we confirmed that T7 RNA polymerase could tolerate an N-terminal fusion, we next developed an assay to quantify mutagenesis via cytidine deamination. We constructed reporter plasmids either having or lacking a T7 promoter sequence upstream of silent drug resistance genes with ACG triplets in place of ATG start codons. In this assay, successful C→T mutagenesis at the *Kan*^{R} or *Tet*^{R} defective start codon yields kanamycin resistant or tetracycline resistant colonies, respectively (**Figure 4.5a**). This assay required optimization because the ACG start codon still seemed to confer some drug resistance in the absence of a mutagen when drug concentrations were low, potentially due to a low level of translation initiation at the ACG codon.[31] However at high drug concentrations, these background colonies disappeared, and colonies only appeared upon mutagenesis with the MP6 plasmid (**Figure 4.5b**).[4]

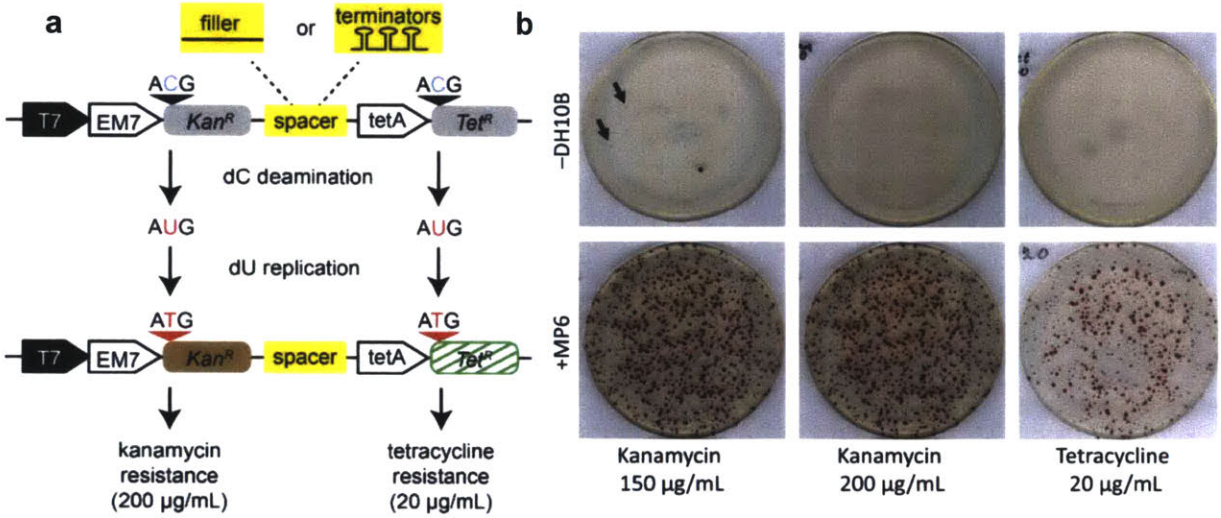

Figure 4.5 Mutation assay overview and optimization.
(**a**) Drug resistance start codon reversion reporter assay for measuring extent of mutagenesis at specific DNA loci. (**b**) At concentrations of less than 200 µg/mL, small colonies (black arrows) appeared on LB + kanamycin + tetrazolium chloride plates with DH10B carrying the reporter plasmid. 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 resulting from MP6 mutagenesis 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 cells carrying the reporter plasmid, while many colonies appeared upon MP6 mutagenesis.

In order to quantify the rate of mutagenesis, we grew cultures from single colonies over defined periods of time. Clonal purification ensures that the starting population does not have any pre-existing mutations and the defined time periods and growth conditions make the assay more reproducible. At the end of the growth period, the cultures were diluted and plated onto LB agar plates with the appropriate antibiotic and tetrazolium chloride, a metabolic dye that stains live cells a dark red color for better contrast during imaging (**Figure 4.6**).[32] After the plates were incubated for 48 hours, they were imaged and colonies were counted using the OpenCFU (3.9.0) software.[33]

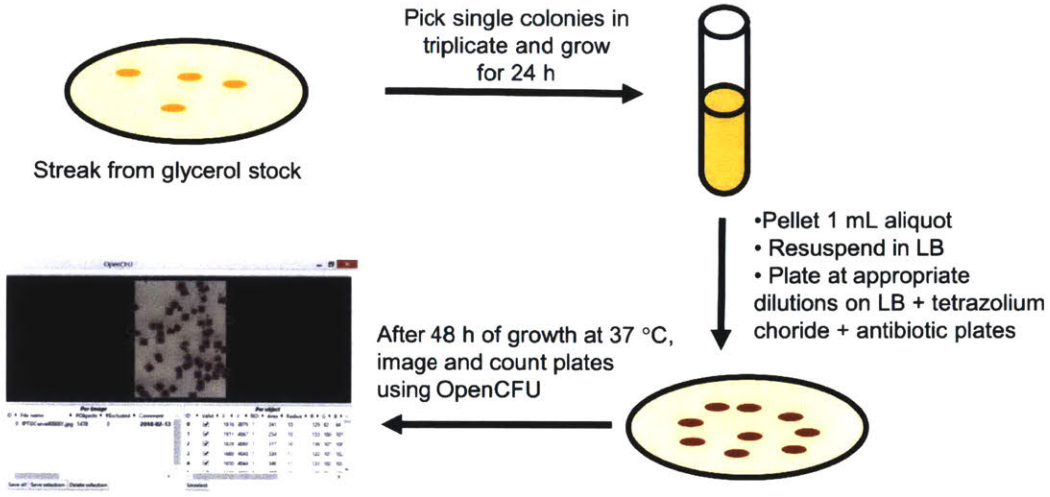

Figure 4.6 Mutation assay workflow.
Glycerol stocks of each sample were streaked on LB-agar with appropriate antibiotics and grown at 37 °C for 24 h to obtain clones. Single colonies were picked in triplicate and grown in LB with appropriate antibiotics and inducers of mutagenesis at 37 °C for 24 h to 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. 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.

Global mutagens such as the MP6 plasmid yielded high levels of kanamycin-resistant colonies regardless of whether the T7 promoter was present in the reporter plasmid or not, consistent with a lack of promoter-based targeting (**Figure 4.7**). In contrast, MutaT7 strains attained significant kanamycin resistance only when reporter plasmids possessed a T7 promoter upstream of the *Kan*^{R} gene (**Figure 4.7** and **Table 4.4**). Expression of a catalytically dead version of MutaT7 (drApo1–T7)[29] yielded kanamycin resistance frequencies similar to background levels, indicating that T7 activity alone was not responsible for the observed increase in kanamycin resistance (**Figure 4.7**).

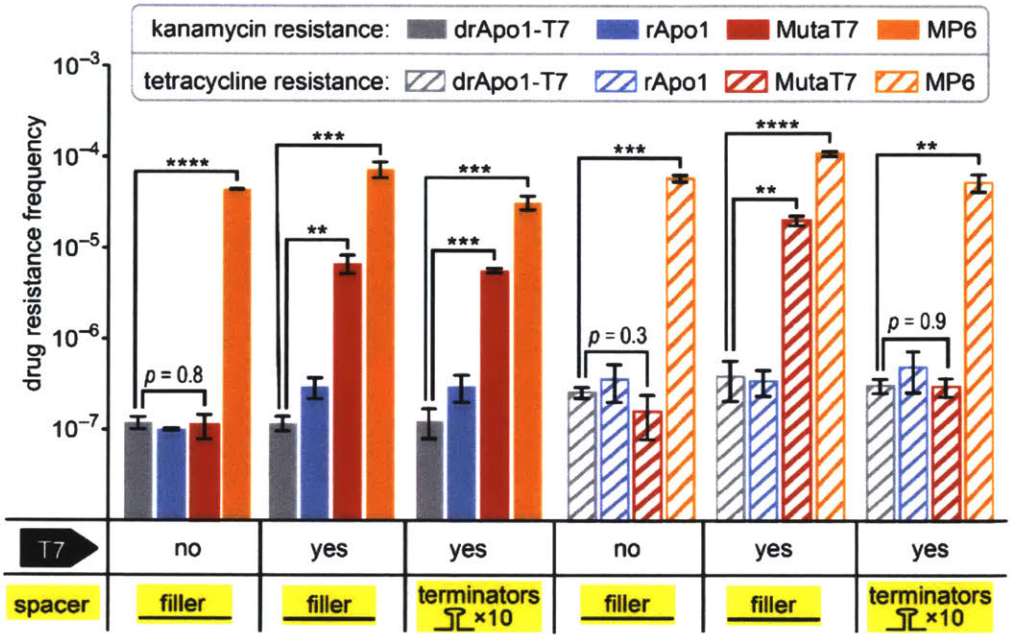

Figure 4.7 MutaT7 introduces mutations downstream of a T7 promoter. Mutagens include deactivated rApo1 fused to T7 RNA polymerase (drApo1–T7; negative control), unfused rApo1 (rApo1), targeted mutagen (MutaT7), and global mutagen (MP6). Values represent mean of independent experiments (n = 3); error 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 shown. See also **Table 4.4**.

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 RNA polymerase is highly processive, we anticipated mutations would also be introduced further downstream of the T7 promoter. We assayed for MutaT7 processivity by inserting a tetracycline-resistance (*Tet*^{R}) gene with an inactive, ACG start codon ~1.6 kb downstream of the *Kan*^{R} gene (**Figure 4.5**). 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 4.7**). Global mutagens again generated tetracycline-resistant colonies at high frequency in all cases, irrespective of the T7 promoter (**Figure 4.7**).

While the T7 promoter demarcates the beginning of the region to be mutated, we also wanted to demarcate the end of the region. To address termination, we used *Kan*^{R}/*Tet*^{R} reporter plasmids in which we separated the silent resistance genes by one or more T7 terminators (**Figure 4.5**). Upon assaying for drug resistance, we found that four copies of the T7 terminator fully constrained mutagenesis to the intended upstream *Kan*^{R} gene (**Figure 4.8**). Using an array of 10 terminators, we observed tetracycline resistance for MutaT7 strains similar to background levels, whereas kanamycin resistance remained high (**Figure 4.7**). Global mutagens again induced high levels of kanamycin- and tetracycline-resistance, irrespective of the terminator array (**Figure 4.7**).

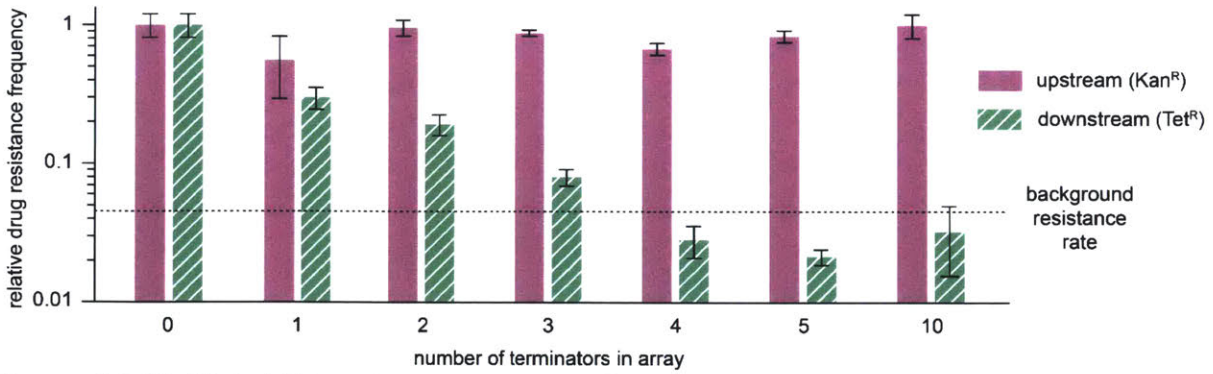

Figure 4.8 Multiple T7 terminators prevent downstream mutations.
After growing the reporter plasmid in the MutaT7 strain (**Table 4.3**) for 24 h, the frequency of kanamycin resistant mutant colonies was relatively constant regardless of the number of terminators between the kanamycin and tetracycline 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; **Table 4.3**).

To further assess whether MutaT7 induces mutagenesis specifically on the target DNA, we evaluated the evolution of resistance to rifampicin[34] and fosfomycin[35]. Resistance can derive from diverse genomic mutations such that the appearance of resistant colonies correlates with off-target mutation rates in the genome,[4,34] analogous to cheating parasites in directed evolution schemes. Selection on either rifampicin- or fosfomycin-treated plates revealed that MutaT7-expressing samples displayed drug resistance frequencies comparable to background. In contrast, high frequencies of antibiotic resistance were observed in all global mutagenesis samples (**Figure 4.9**).

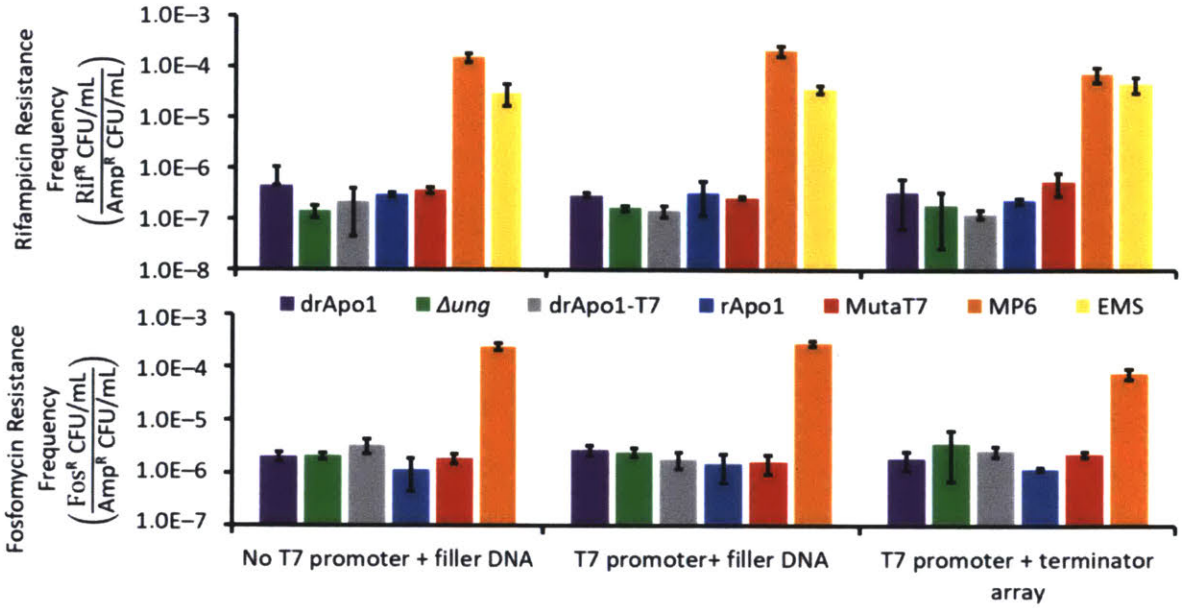

Figure 4.9 Off-target mutagenesis in the *E. coli* genome.
Rifampicin and fosfomycin resistance frequency data show a high mutagenesis rate only in the presence of MP6 or ethyl methanesulfonate (EMS), suggesting that neither MutaT7 nor the negative controls (**Table 4.3**) mutagenize the *E. coli* genome appreciably. Values represent mean of independent experiments (n = 3); error bars represent one standard deviation. See also **Table 4.4**.

In theory, another benefit of targeted mutagenesis is the capacity to attain much larger library sizes by avoiding toxic mutations in essential, off-target genes. On the basis of 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,[4] we observed very low viability in all populations treated with global mutagens, such as MP6 or EMS. In contrast, populations expressing MutaT7 possessed viability similar to untreated cells (**Figure 4.10a**). We also found that the total number of kanamycin-resistant colonies was similar between MutaT7 and globally mutagenized samples (**Figure 4.10b**) despite the somewhat lower mutagenesis rate of the MutaT7 construct relative to MP6 (**Figure 4.7**; the average kanamycin resistance frequency for MutaT7 was 6.7×10^{-6} versus 5.7×10^{-5} for MP6). This observation highlights that the use of MutaT7 to maximize on-target mutations while simultaneously minimizing off-target mutations results in larger productive library sizes.

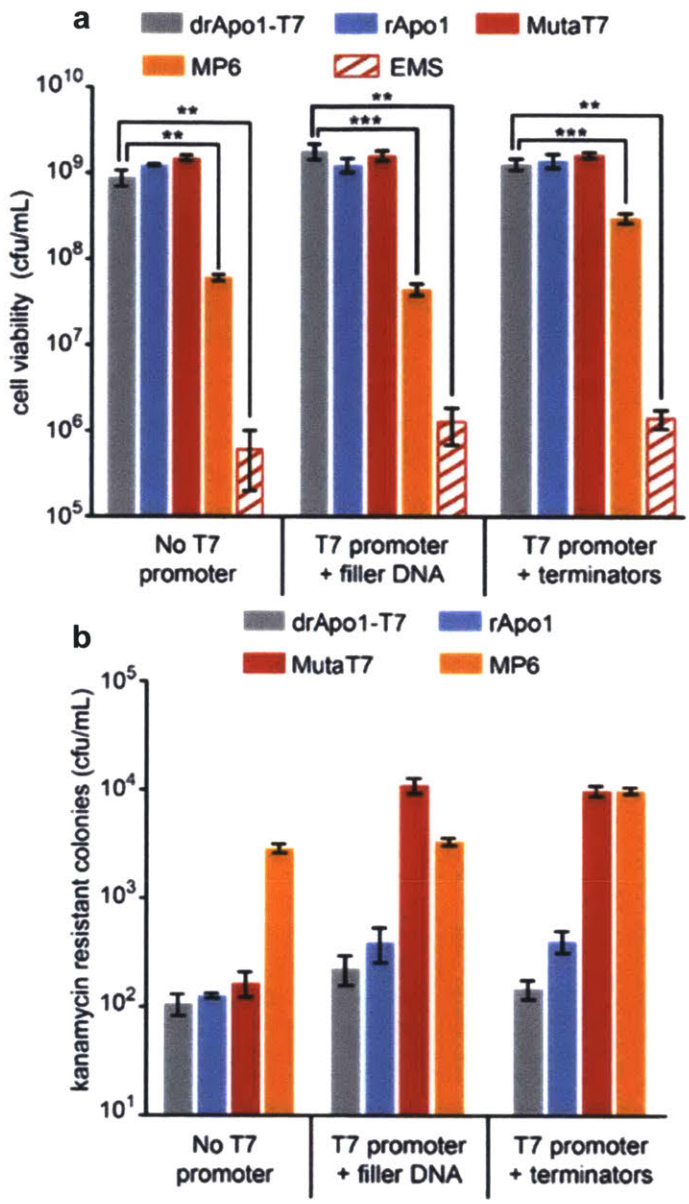

Figure 4.10 MutaT7 has minimal impact on cell viability.
(**a**) Viability data for cell populations in **Figure 4.2** along with drApo1-T7 populations treated with EMS, as determined by the number of Amp^R resistant colonies. (**b**) Total number of kanamycin resistant colonies for populations in **Figure 4.2**. 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 shown. See also **Table 4.4**.

4.3.3 Mutational accumulation and evolution with MutaT7

We next attempted to use MutaT7 to evolve ectopically expressed *folA* gene variants that confer trimethoprim resistance. The *folA* gene encodes dihydrofolate reductase, the direct target of trimethoprim, but *folA* mutations are just one of many potential routes to trimethoprim resistance.[36] We used either global mutagenesis or MutaT7 to mutagenize *E. coli* carrying a T7-targeted episomal copy of *folA*. We then Sanger-sequenced colonies that grew on trimethoprim plates. We observed that 29 of 44 trimethoprim-resistant colonies mutagenized using MutaT7 had a mutation known to confer resistance[37] in the episomal *folA* promoter (**Table 4.1**, **Figure 4.11**). In contrast, none of the 43 trimethoprim-resistant colonies obtained using the global mutagen contained mutations in the episomal *folA* gene. Instead, they presumably gained trimethoprim resistance via undesired mutations in the *E. coli* genome. The ability of MutaT7 to generate a high rate of true positives in the desired episomal gene target, whereas global mutagenesis exclusively generated cheaters (false positives), highlights a key advantage of MutaT7.

Table 4.1 Sanger sequencing data for Figure 4.11.

Mutagen	Biological Replicate	Total Colonies Sequenced	On-target folA Mutation (plasmid)	Off-target Mutation	Relative On-target Frequency
MutaT7	1	14	11	3	78.57%
	2	15	8	7	53.33%
	3	15	10	4	66.67%
	Total	44	29	15	65.91%
MP6	1	15	0	15	0%
	2	12	0	12	0%
	3	16	0	16	0%
	Total	43	0	43	0%

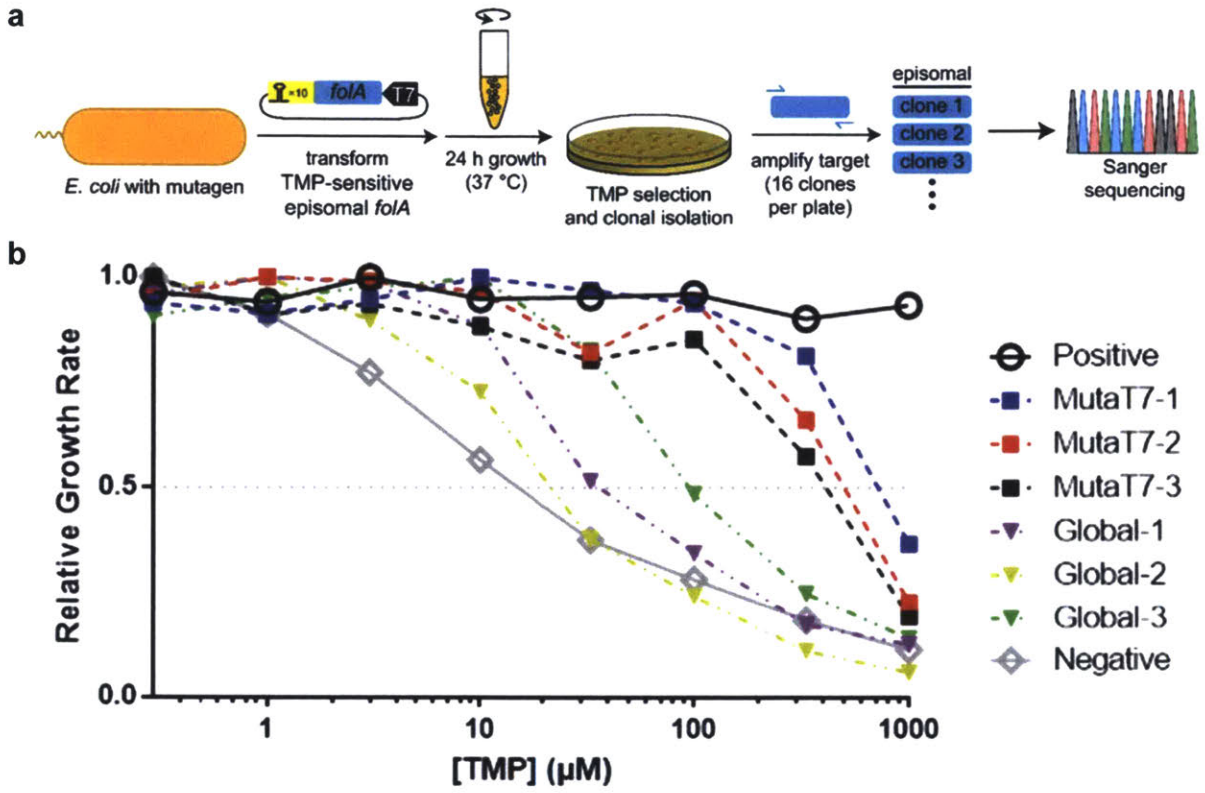

Figure 4.11 Directed evolution of *folA* using MutaT7 results in a much lower false positive frequency than does a global mutagen.
(**a**) Schematic of a directed evolution experiment on *folA* (promoter and protein coding sequence of dihydrofolate reductase from E.coli) designed to measure the frequency of true and false positives following mutagenesis and selection with trimethoprim (TMP). Clones propagating an episome with wild-type *folA* downstream of a T7 promoter were mutagenized with MutaT7 or a global mutagen (MP6) in the absence of selection pressure. Selection on LB-agar plates with TMP enabled isolation of TMP-resistant colonies. Subsequent amplification and Sanger sequencing of episomal *folA* genes was used to assess the frequency of true positives (drug-resistant mutations in episomal *folA*) and false positives (drug-resistant mutations somewhere else in genome). (**b**) Summary of bacterial growth curve data measuring extent of TMP resistance in evolved isolates. Growth rates in response to increasing concentrations of TMP were determined for a representative isolate from each biological replicate along with a positive control (episomal *folA*, but with a strong promoter instead of wild-type promoter) and a negative control (drApo-T7 with episomal *folA*). After determining maximal growth rate within each sample, growth rates were normalized to the highest rate within each sample series, yielding the relative growth rate (y-axis) at each TMP concentration (x-axis).

We next turned to DNA sequencing to better understand the distribution of mutations introduced by MutaT7 beyond the defective ACG codons used in the codon reversion assays. We allowed an *E. coli* population expressing MutaT7 and the episomal Kan^R/Tet^R reporter plasmid to drift in the absence of selection pressure for 15 days prior to isolation of episomal DNA from clones (**Figure 4.12a**). Sanger sequencing of the target

episomal region revealed mutations at multiple loci throughout the *Kan*^{R} target gene, independent of selection pressure (**Figure 4.12b**).

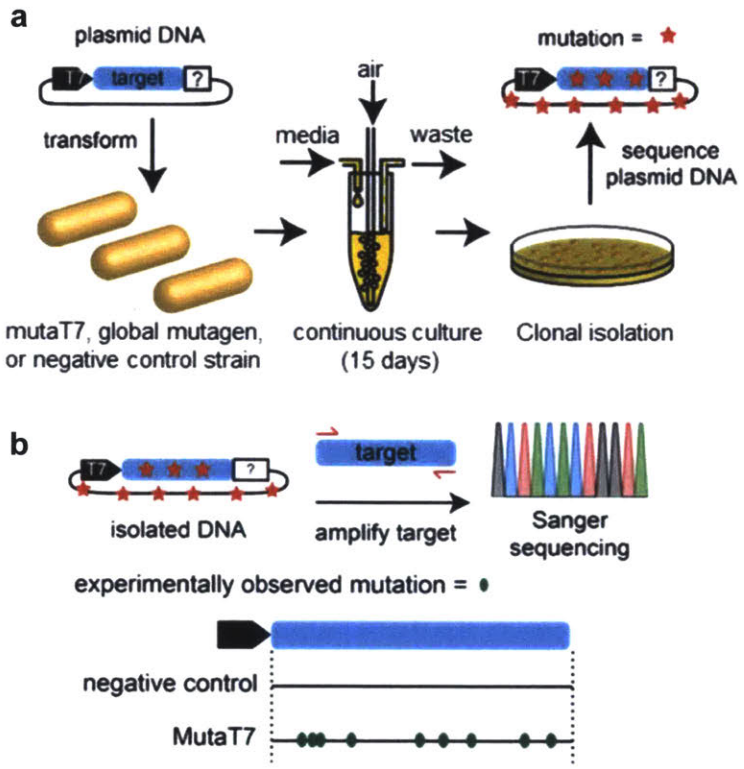

Figure 4.12 MutaT7 introduces mutations throughout the target region.
(**a**) Reporter construct and continuous culture experiment to assess mutation accumulation under drift conditions. (**b**) Schematic and representation of mutations observed by Sanger sequencing 96 clones in the indicated cell populations following 15 days of continuous growth in the absence of selection pressure.

Next, we employed Illumina sequencing to identify mutations anywhere in the episomal reporter DNA sequence obtained from clones of the *E. coli* populations in **Figure 4.12a**. This experiment assesses on- versus off-target mutagenesis across a ~10 kb stretch of DNA containing only ~1 kb of intended target DNA. MutaT7 samples displayed many mutations throughout the episome when the terminator array was removed but the T7 promoter was maintained (**Figure 4.13a**). Treatment with the MP6 global mutagen also led to mutations throughout the entire episomal DNA. In contrast, mutations in MutaT7 strains 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 4.13a**). 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 4.13b**, red striped bars).

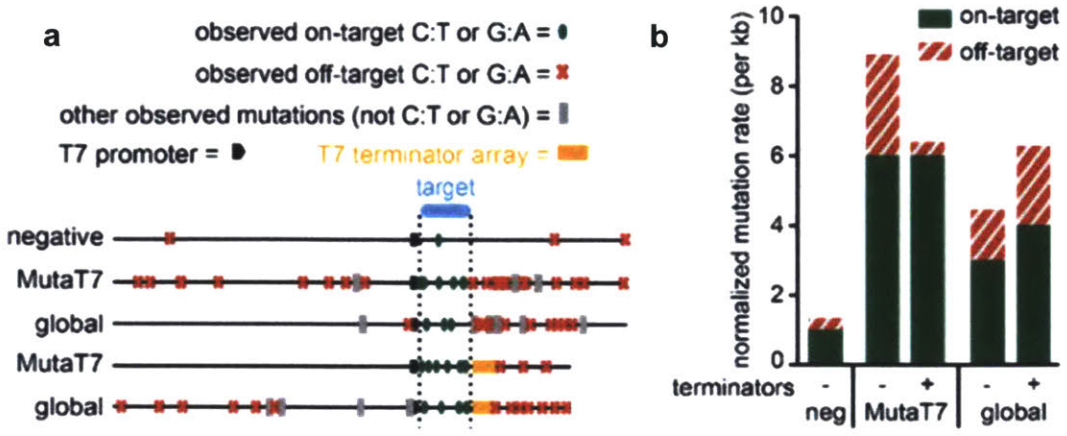

Figure 4.13 Off-target:on-target ratio by next-generation sequencing.
(**a**) On-target (green oval) and off-target (red x) mutations identified by sequencing episomes propagated in the presence of targeted (MutaT7) and global (MP6) mutagens. (**b**) Normalized mutation frequency (number of mutations observed divided by kb of DNA sequenced in associated regions) for data in panel **a**.

In order to observe more mutations that are widely distributed within a reasonable time frame, we employed counterselection rather than mutational drift. MutaT7 was targeted to an episomal *rpsL* allele that confers streptomycin sensitivity (**Figure 4.14a**). Sanger sequencing of streptomycin-resistant mutants isolated from a MutaT7-expressing strain of *E. coli* again revealed that multiple mutations appeared throughout the targeted *rpsL* gene, with ~90% C→T mutations and ~10% G→A mutations (**Figure 4.14b–c**). The bias for C→T mutations likely arises from the greater accessibility of the coding DNA strand for cytidine deamination. The G→A mutations likely result from less frequent cytidine deamination of the template, non-coding strand, which is partly buried in the active site of T7 RNA polymerase during transcription.[38] MP6 on the other hand was able to variety of transitions and transversions that are not possible via cytidine deamination alone (**Figure 4.14b–c**).

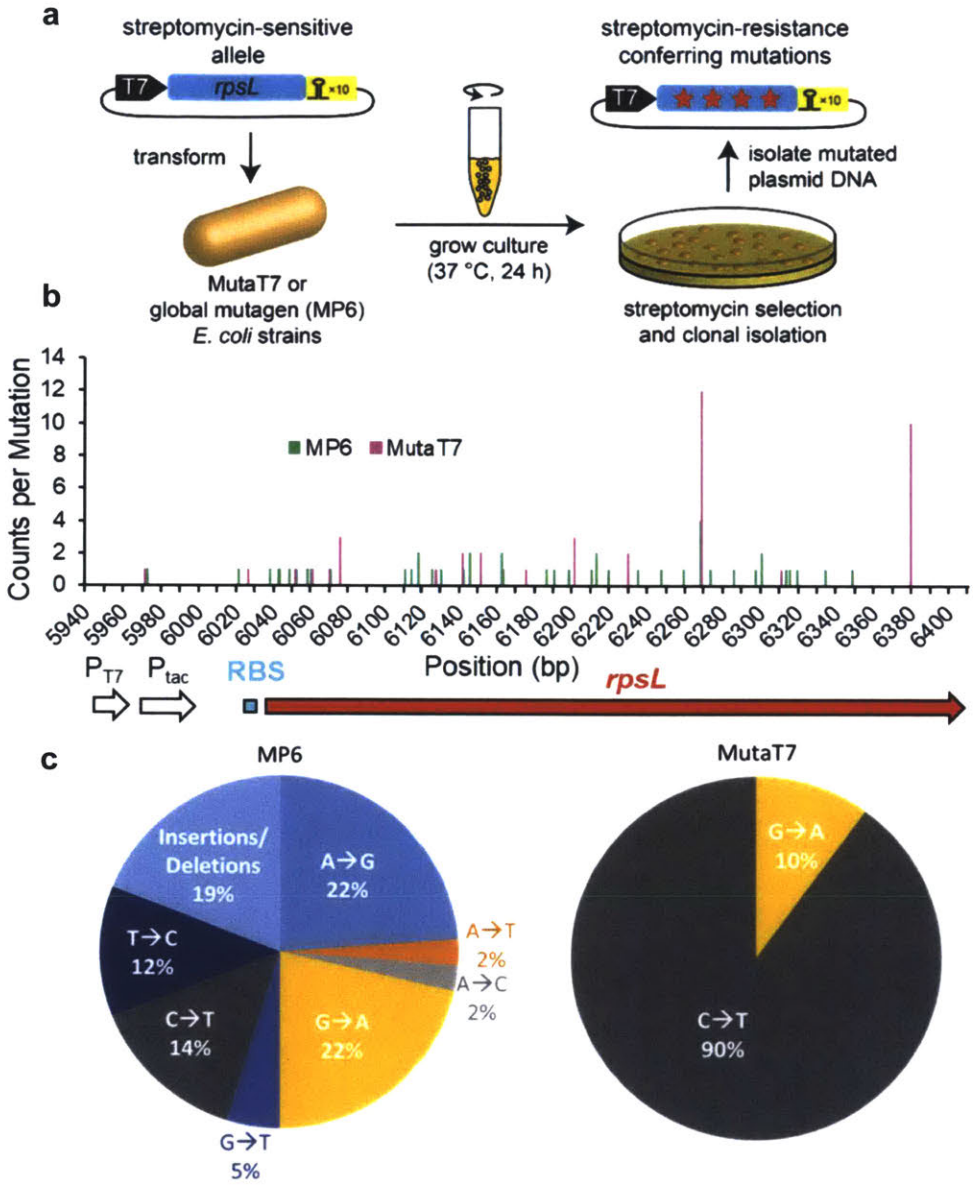

Figure 4.14 Mutational spectra of MP6 and MutaT7.
(**a**) Schematic of streptomycin resistance counter-selection assay, which is designed to enrich for mutations that nullify streptomycin sensitivity. Such sensitivity is initially conferred by a streptomycin–sensitive allele of *rpsL* downstream of a T7 promoter on a reporter plasmid. (**b**) The position of various mutations throughout the T7 promoter + rpsL reporter plasmid determined by Sanger sequencing of 48 streptomycin resistant mutants from the MP6 strain and 42 streptomycin resistant mutants from the MutaT7 strain. (**c**) Distribution of types of mutations in panel **b**.

Two current disadvantages of MutaT7 are its limited mutational spectrum and an apparent strand bias for the coding strand when using a single T7 promoter. We hypothesized that we could eliminate the strand bias by introducing a second T7 promoter that would recruit MutaT7 to the 3'-end of the target DNA and enable processive activity in the opposing direction. Indeed, we found that installing an additional antisense T7

promoter led to the accumulation of both G→A and C→T mutations throughout the target gene during continuous culturing (**Figure 4.15a–b**). Furthermore, the average number and range of mutations per clone increased over time (**Figure 4.15c**). The latter observation indicates that, in contrast to global mutagenesis methods where the organism often rapidly silences mutagen expression, the high on-target to off-target mutation ratio of MutaT7 enabled long-term maintenance of mutagen expression in cells.

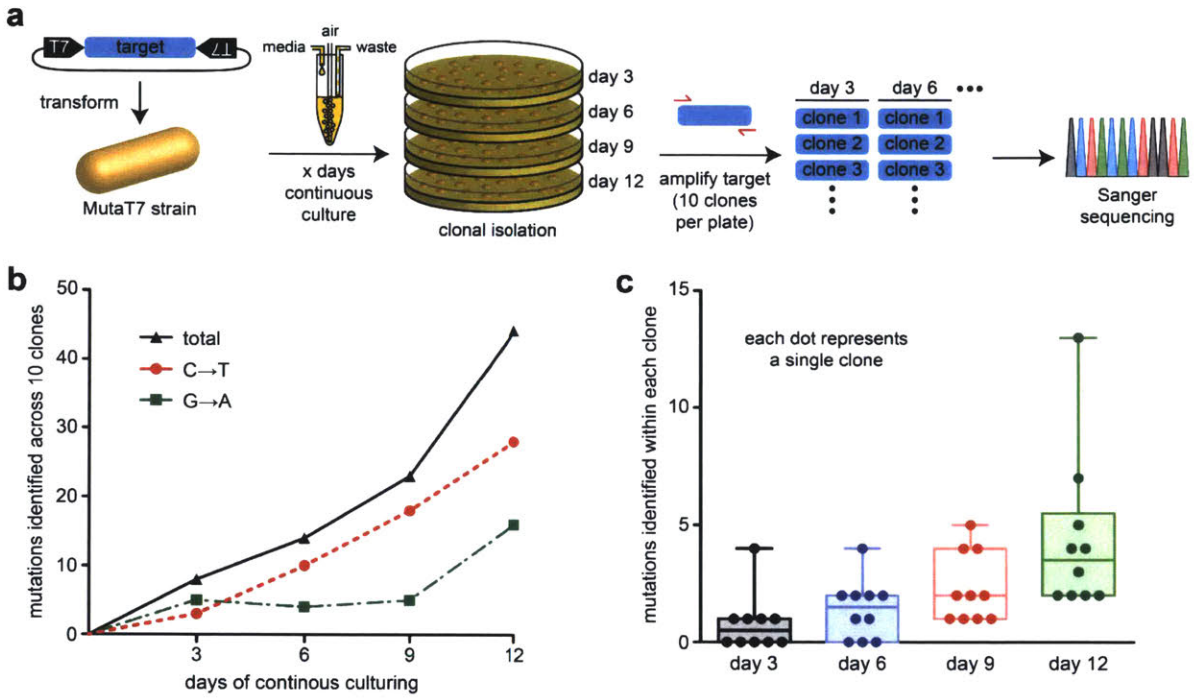

Figure 4.15 Dual T7 promoters introduce mutations in both strands.
(**a**) Diagram of continuous culture conditions used to propagate a dual promoter episome in cells expressing MutaT7, along with details for downstream Sanger sequencing analysis. (**b**) Graphic of mutations observed by Sanger sequencing 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). (**c**) Box and whisker plot of mutations from panel **b**, 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.

4.3.4 Expanding the mutational spectrum with adenosine deaminase

To further expand MutaT7's mutational spectrum, we turned to a recently developed DNA damaging enzyme with adenosine deaminase activity, TadA, which has been fused to dCas9 to successfully induce A→G mutations at specific DNA loci.[11] We replaced rApo1 in the MutaT7 strain (**Table 4.3**) with tadA, either keeping the existing Gly-Ser linker or replacing it with the more rigid XTEN linker.[39] We also constructed a new series of reporter plasmids that have *Kan*^{R} and *Tet*^{R} premature TAG stop codons at sites that originally coded for TGG. Successful mutagenesis by tadA converts the TAG stop codon back into a TGG tryptophan codon. When we performed the mutagenesis assay (**Figure 4.6**) with the new tadA strains (**Table 4.3**) and reporter plasmids, we observed that the tadA–T7 fusions introduced more mutations upon IPTG induction than did tadA alone. Furthermore, significant mutagenesis above background only occurred when the reporter had a T7 promoter (**Figure 4.16**). We note that both tadA–T7 fusions appeared to have less mutagenic activity than the rApo1–T7 fusion (**Figure 4.7**), producing a lower frequency of kanamycin and tetracycline resistant mutants. It was also necessary to induce the $P_{A1lacO\text{-}Tenth}$ promoter with IPTG and thus express higher levels of the tadA–T7 fusion. In contrast, the mutagenesis experiments with the rApo1–T7 fusion were performed with the $P_{A1LacO\text{-}Tenth}$ promoter in the repressed state, suggesting that very low expression was required to achieve mutagenesis and that the fusion protein was thus highly active.

Expression of the tadA–T7 fusions largely did not result in a greater number of rifampicin resistant mutants, suggesting low off-target activity (**Figure 4.17a**). The tadA–T7 fusions also did not appear to generally impact cellular viability (**Figure 4.17b**). However, we note that, specifically when the reporter plasmid had a T7 promoter and was induced with IPTG, the tadA-GGS-T7 fusion displayed a higher frequency of rifampicin resistant mutants and a lower number of ampicillin resistant colonies. A likely explanation for this unexpected result is that the transcription activity of tadA-GGS-T7 was interfering with reporter plasmid replication, since the presence of a T7 promoter should not affect the overall toxicity and off-target activity of the tadA-GGS-T7 fusion protein. T7 RNA polymerase read-through of the origin of replication has indeed been shown to result in a greater amount of plasmid loss.[40] While cell death due to mutagenic toxicity can result in fewer colonies in our assay, the loss of the plasmid that confers ampicillin resistance can also contribute. In the latter case, the elevated rifampicin resistance frequency we observed would simply be an artifact of the plasmid loss, since the number of surviving cells would actually be undercounted and the number of rifampicin resistant colonies counted would be accurate.

Another implication of the putative plasmid loss in the tadA-GGS-T7 +T7 promoter +IPTG condition (**Figure 4.17**) is that the tadA-GGS-T7 fusion had greater transcriptional activity and/or a higher expression level than the tadA-XTEN-T7 fusion, which did not appear to cause plasmid loss but still mutagenized the reporter plasmid. While an undercount of the number of surviving cells in the tadA-GGS-T7 +T7 promoter +IPTG condition would lead to an overestimate of rifampicin resistance frequency, the on-target mutation frequency measurement should be accurate in **Figure 4.16** since the mutated plasmid is required for the acquisition of kanamycin or tetracycline resistance. In future assays, we will quantify plasmid loss by measuring the ratio of ampicillin-resistant colonies (the number of plasmid-containing cells) to streptomycin resistant colonies (the

true number of surviving cells), and thus avoid overestimating off-target mutagenesis rates. While the tadA-XTEN-T7 fusion seemed less transcriptionally active, it produced the same on-target mutation frequencies as tadA-GGS-T7, suggesting that a high level of T7 RNA polymerase transcription is unnecessary and perhaps undesirable for achieving mutagenesis.

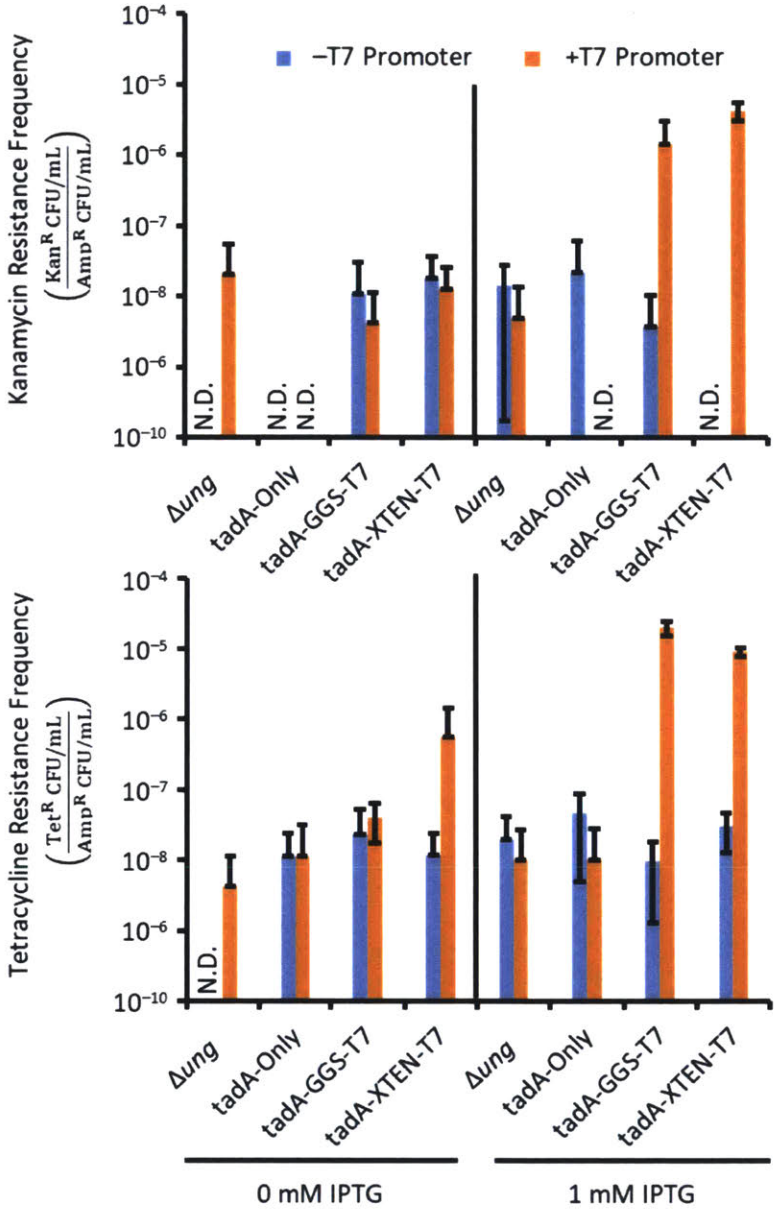

Figure 4.16 TadA–T7 fusions introduce A→G mutations in targeted manner.

Mutagens include tadA alone (tadA-Only), tadA fused to T7 RNA polymerase through a Gly-Ser linker (tadA-GGS-T7), and tadA fused to T7 RNA polymerase through an XTEN linker (tadA-XTEN-T7). The Δ*ung* strain was used as a negative control. Mutagenesis rates were measured with or without IPTG induction. Values represent mean of independent experiments (n = 3); error bars represent standard deviation. N.D. = not detected.

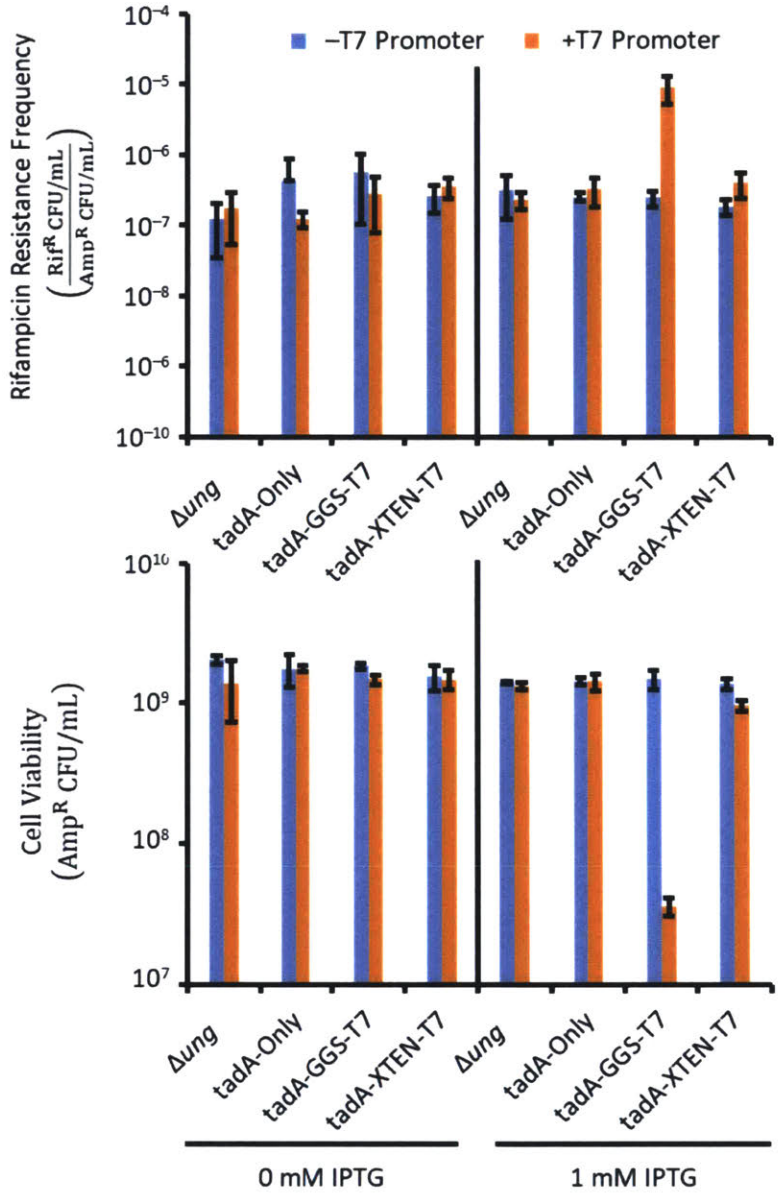

Figure 4.17 TadA–T7 fusions have low off-target activity and toxicity.

(**a**) Rifampicin resistance frequency data for cell populations in **Figure 4.16** suggests that tadA alone and the tadA–T7 fusions did not introduce significant off-target mutations in most cases. (**b**) Viability data for cell populations in **Figure 4.16**, as determined by the number of Amp^R resistant colonies, suggests the tadA–T7 fusions were not significantly toxic. Values represent mean of independent experiments (n = 3); error bars represent standard deviation.

4.3.5 Inhibiting DNA repair with ugi rather than a *ung* deletion

Finally, we note that repair of deoxyuridine must be prevented to observe significant mutagenesis with the rApo1–T7 RNA polymerase fusion. Impairment of DNA repair has also been shown to be important for other cytidine deaminase-based systems.[4,8] Although we used Δ*ung* cells to address this issue in the aforementioned experiments, a more flexible alternative is to co-express MutaT7 with the uracil glycosylase inhibitor (ugi), a protein that can inhibit UNG activity in many prokaryotes and eukaryotes.[4,8,41] Such co-expression resulted in a high rate of mutagenesis similar to that achieved using Δ*ung* cells (**Figure 4.18**). Ugi thus eliminates the need to delete *ung* to achieve efficient mutagenesis with MutaT7, significantly increasing the flexibility of our system.

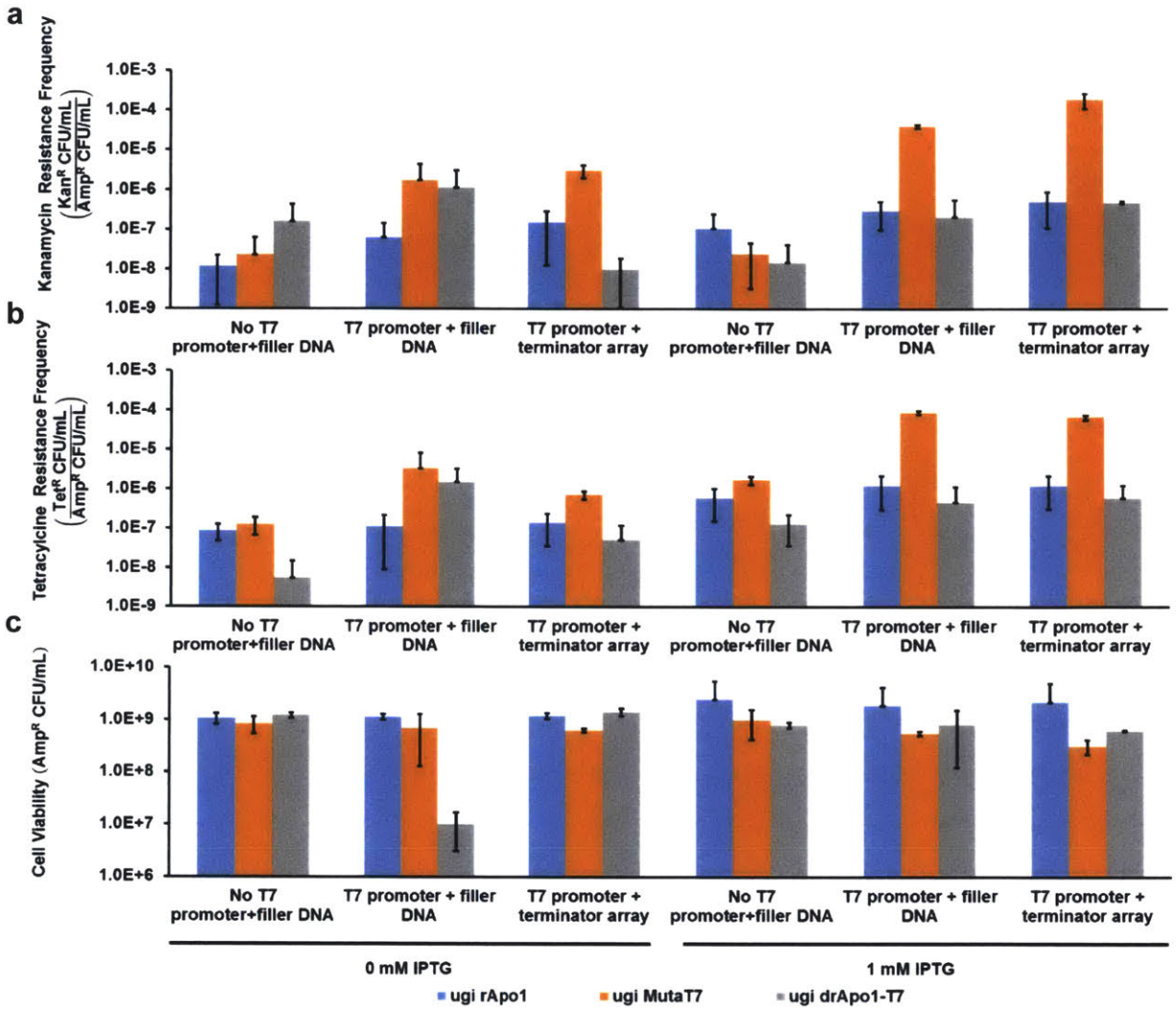

Figure 4.18 Ugi expression increases mutagenesis by inhibiting dU→dC repair.
(**a**) Kanamycin resistance frequency data for the ugi rApo1 and ugi drApo1–T7 negative control strains and the ugi MutaT7 mutagenic strain (**Table 4.3**) with various reporter plasmids showed that the ugi protein can increase mutagenesis when expression of ugi and MutaT7 from the $P_{A1lacO\text{-}Tenth}$ promoter is induced with IPTG. (**b**) Tetracycline resistance frequency data for the same experiment performed in panel **a**. (**c**) Cell viability as determined by the number of ampicillin resistance colonies for the same experiment performed in panel **a**.

4.4 Discussion

In summary, our results show that DNA damaging enzymes can be targeted by the action of T7 RNA polymerase transcription between a T7 promoter and a terminator array. Presumably, the T7 RNA polymerase allows for a high rate of on-target mutagenesis by localizing the DNA damaging agent to the region of interest, and by generating the requisite single-stranded DNA substrate upon which rApo1 acts. Our data suggest that only a minimal level of MutaT7 expression is required to achieve high rates of mutagenesis, and that increasing the expression would in fact not be beneficial owing to interference with plasmid replication and toxicity. In all of our experiments, mutations were targeted to a single T7 promoter on a single-copy BAC episome, so it is possible that the T7 promoter at the target site may become saturated at very low levels of MutaT7 expression, with any additional MutaT7 expression only contributing to potential off-target activity. Slightly higher MutaT7 activity may be beneficial for targets with high copy numbers, such as plasmids with the pUC origin, but this hypothesis remains to be explored.

The discrepancy between the mutagenesis rates of the rApo1 fusion and the tadA fusion suggests that the activity of the DNA damaging enzyme is the main determinant of the mutagenesis rate rather than the activity of the T7 RNA polymerase. Furthermore, in the hypothetical case that a single-copy T7 promoter is saturated at even low MutaT7 expression levels, the only way to increase on-target mutagenesis would be to introduce more T7 promoters at the target site, or to increase the activity of the DNA damaging enzyme. Variants of rApo1 and another cytidine deaminase, CDA1, have recently been engineered to have higher activity and reduced sequence context preferences,[42] and could be quite useful in boosting the mutagenesis rate of MutaT7 even further.

Since the rApo1–T7 or tadA–T7 fusions can introduce C→T or A→G mutations, respectively, on either the coding or template strand, the mutational spectrum of MutaT7 is able to make a wide range of mutations in amino acid space during a single round of mutagenesis. A single C→T change in the coding strand can result in 11 possible missense mutations and 2 possible nonsense mutations. A single G→A mutation in the coding strand (by introducing a C→T mutation in the template strand) can result in 14 possible missense mutations and 1 possible nonsense mutation that are not redundant with C→T mutations in the coding strand. Therefore, rApo1 can make 28 possible mutations at the amino acid level. Adding A→G and T→C mutations from tadA doubles the number of accessible non-redundant mutations to 56 that are accessible by a single nucleotide change, including mutations that convert stop codons to a non-stop codon. Thus, we are able to access a third of the 174 possible amino acid substitutions that can theoretically be attained by a single nucleotide change during a single round of mutagenesis in the standard genetic code.[43] Additional amino acid changes can be achieved in successive rounds of mutagenesis. To further expand the number of possible amino acid substitutions in a single round of mutagenesis, we are now exploring the use of other DNA damaging enzymes that make other types of mutations or that have broad, non-specific activity. Unfortunately, there are few enzymes beyond cytidine deaminases that can induce mutagenic damage in DNA specifically. However, there exists a vast array of RNA modifying enzymes[44] that could potentially be re-engineered to recognize DNA. For example, the tadA adenosine deaminase used here was derived from an RNA adenosine deaminase by engineering it to recognize DNA.[11]

Another major goal for future work with MutaT7 will be to apply it to organisms other than *E. coli*. We showed that ugi, a protein inhibitor of ung, can be co-expressed with MutaT7, and thus obviate the need to delete ung as a prerequisite for targeted mutagenesis. Although ugi was found in the PBS2 bacteriophage of *Bacillus subtilis*, it has remarkably broad activity and inhibits ung homologs in vastly different organisms, including *E. coli* and humans.[45] Similarly, the T7 RNA polymerase has been previously shown to function in a wide array of organisms, including human cells,[46] *Bacillus subtilis*,[47] tobacco plants,[48] CHO cells,[49] *Vibrio natriegens*,[50] and yeast.[51] After designing appropriate expression vectors, we anticipate that MutaT7 can function in any of these organisms. We also plan to apply MutaT7 in our adenovirus-mediated, human cell-based continuous directed evolution platform (Chapter 3).

4.5 Conclusion

In summary, the processively acting MutaT7 chimera can selectively direct mutations to large, yet well-defined, regions of DNA *in vivo*. We show that both cytidine deaminases and adenosine deaminases[11] can be used to enable the creation of rich and diverse DNA libraries *in vivo*. The ubiquitous applicability of T7 RNA polymerase in diverse organisms[48-51] suggests that MutaT7 will prove useful in a broad range of evolutionary and synthetic biology settings.

4.6 Materials and Methods

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. Plasmid sequences can be in our previous work.[52] pCMV-ABE7.10 was a gift from David Liu (Addgene plasmid #102919).[11]

Reagents: The following reagents were obtained as indicated: Kanamycin monosulfate, fosfomycin, agar, and chloramphenicol (Alfa Aesar J61272, J66602, A10752, and B20841, respectively); tetracycline hydrochloride (CalBioChem 58346); rifampicin (TCI R0079); ampicillin (Fisher Bioreagents BP1760-25); streptomycin sulfate (MP Biomedical 100556); tetrazolium chloride, L-rhamnose, antifoam-204, and ethyl methanesulfonate (Sigma-Aldrich T8877, W373011, A8311, and M0880, respectively); L-arabinose and cycloheximide (Chem-Impex 01654 and 00083, respectively); and lysogeny broth (LB; Difco 244620); anhydrous sodium phosphate dibasic and monobasic sodium phosphate (Mallinckrodt 7917 and 7892, respectively); potassium chloride and isopropyl β-D-1-thiogalactopyranoside (Sigma P9333 and I6758, respectively); magnesium sulfate (Macron 6070-12); *o*-Nitrophenyl-β-galactoside and egg-white lysozyme (VWR 0789 and 0663, respectively); PopCulture lysis reagent (EMD Millipore 71092-4); 2-mercaptoethanol (Bio-Rad 161-0710); trimethoprim (Matrix Scientific 058373).

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[8] and the T7 RNA polymerase gene was amplified from pTara.[53] Mutation assay reporter plasmids utilizing the single-copy BAC origin and the terminator arrays of the UUCG-T7 derivative of the T7 terminator[40] were generated by serial insertion of the annealed oligos NheI-UUCG-BamHI S and NheI-UUCG-BamHI AS (**Table 4.5**). The *folA* gene was amplified from DH10B genomic DNA. All *E. coli* strains used in this work were engineered using lambda red recombineering strategies described in detail below.

Mutation assay: To assess mutagenesis rates, the control (Δ*ung*, rApo1, drApo1, and drApo1–T7; **Table 4.3**) and mutagenic strains (MutaT7 and MP6; **Table 4.3**) ($Strep^R$) 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 L-arabinose (with 10 μg/mL chloramphenicol for the MP6 strain, **Table 4.3**), 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 L-arabinose. Each resuspension was plated on LB agar plates with 50 μg/mL tetrazolium chloride (a metabolic contrast dye for visualizing colonies)[32] 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 (**Table 4.4** and **Figure 4.10**).
- 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 (**Table 4.4** and **Figure 4.10**). 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 (**Table 4.4** and **Figure 4.7**).
- 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 (**Table 4.4**). 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 (**Table 4.4** and **Figure 4.7**).
- 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 (**Table 4.4**). 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 (**Table 4.4** and **Figure 4.9**).
- 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 (**Table 4.4**). 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 (**Table 4.4** and **Figure 4.9**).

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.[32] The colonies were then counted using the software OpenCFU (3.9.0),[33] with the minimum colony radius set to 3, the maximum colony radius set to 50, and the regular threshold set to 4.

The same assay as above was also used to assess the mutation rate of the ugi rApo1, ugi MutaT7, and ugi drApo1–T7 strains (**Table 4.3**), except that instead of L-arabinose, either 0 mM or 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the liquid overnight cultures as a control or to induce mutagenesis, respectively.

The same assay as above was also used to assess the mutation rate of the tadA-Only, tadA-GGS-T7, and tadA-XTEN-T7 strains (**Table 4.3**), except that instead of L-arabinose, either 0 mM or 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the liquid overnight cultures as a control or to induce mutagenesis, respectively. Also, the BAC-Kanstop-Tetstop or BAC-T7-Kanstop-Tetstop reporter plasmids were used (see Appendix B.2 for sequences).

Chemical mutagenesis with ethyl methanesulfonate (EMS): Mutagenesis with EMS was performed as previously described.[1] 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 resistant colonies were counted after EMS treatment in CFU/mL to measure the viability after mutagen treatment (**Table 4.4** and **Figure 4.10**). 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 (**Table 4.4** and **Figure 4.9**).

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 (**Table 4.3**) 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 (**Table 4.5**).

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 (**Table 4.3**) in 20 mL cultures using a previously described multiplex bioreactor setup.[54] 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 L-arabinose (+10 μg/mL chloramphenicol in the case of the MP6 strain (**Table 4.3**)) 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.

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 (v. 0.7.12-r1039),[55] with flag –t 16, and sorted and indexed bam files were generated using samtools (v 1.3).[56] These bam files were processed using samtools mpileup with flags –excl-flags 2052, -d 10000000 and the same plasmid reference sequences used for mapping[57]. Read coverages and alleles counts and frequencies were tabulated at each position of the reference sequence in each sample for down-stream analysis. 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 (**Table 4.3**) after 15 d of continuous culture carrying the T7 promoter + terminators reporter plasmid. The primers 2165 and 1197 (**Table 4.5**) were used to amplify and Sanger sequence the Kan^R gene.

Lambda *red* recombineering: The *E. coli* genome was edited using seamless lambda red recombineering with ccdB counterselection, as previously described.[58] 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 L-arabinose and L-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 msec, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in super optimal broth with catabolite repression (SOC) with 2 mg/mL L-arabinose at 30 °C for 2 h, then plated on LB agar plates with 50 μg/mL kanamycin and 2 mg/mL L-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 L-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 L-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 L-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 L-arabinose at 30C for 2 h, and then were washed once with LB to remove the L-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 (**Table 4.3**) were generated using the primers in **Table 4.5**.

The following modifications (**Table 4.2**) were made in the following order to obtain the strains used in this work:

Table 4.2 Modifications made to the *E. coli* genome.

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-rApoI 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.[59]
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.[60]
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.[25]
Replacement of promoter BBa_J23114 with $P_{A1lacO\text{-}Tenth}$	Δ(*araA-leu*)*7697*::[$P_{A1lacO\text{-}Tenth}$ *rApo1* or *MutaT7*]	5'-prApoI::kanccdB and 3'-prApoI::kanccdB	PA1lacO-1 F and PA1lacO-1 R used to amplify from pA1lacO-tenth gene block (**Table 4.5**)	The BBa_J23114 promoter from the Anderson Collection[61] that controlled the expression of rApo1 or MutaT7 from the DH10B genome was replaced with the tightly-repressed promoter $P_{A1lacO\text{-}Tenth}$,which is a weaker version of the P_{A1lacO} promoter.[62]
Deactivation of rApoI	Δ(*araA-leu*)*7697*::[$P_{A1lacO\text{-}Tenth}$ d*rApo1* or *drApo1-T7*]	5'-drApoI::kanccdB and 3'-drApoI::kanccdB	drApoI S and drApoI AS (annealed oligos)	The E63Q mutant of rApo1 cytidine deaminase has been shown to be catalytically inactive.[29] E63Q mutant negative control strains were made.
Reinsertion of ung	*ung*$^{+}$	5'-Ung kanccdB and 3'-Ung kanccdB	Ung-Forward and Ung-Reverse used to amplify from DH10B genomic DNA	Used to reverse the *ung* deletion and generate ung+ control strains.
Insertion of ugi	*ugi*	5'-ugi kanccdB and 3'-ugi kanccdB	Ugi-Forward and Ugi-Reverse used to amplify from MP6	Ugi is a small protein that inhibits uracil DNA glycosylase (ung)[4]. Ugi was inserted under the $P_{A1lacO\text{-}Tenth}$ promoter preceding MutaT7 as an alternate strategy for minimizing dU→dC repair.

Insertion of tadA only	Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *tadA*]	5'-drApol::kanccdB and 3'-drApol::kanccdB	tadA-Forward and tadA-Only-Reverse used to amplify from pCMV-ABE7.10	Replacement of MutaT7 in the MutaT7 strain with tadA, an adenosine deaminase.[11] Note, only a single monomer of tadA* was incorporated.
Insertion of tadA-XTEN	Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *tadA-XTEN-T7*]	5'-drApol::kanccdB and 3'-drApol::kanccdB	tadA-Forward and tadA-XTEN-Reverse used to amplify from pCMV-ABE7.10	Replacement of rApo1-GGS in the MutaT7 strain with tadA-XTEN, an adenosine deaminase.[11] Note, only a single monomer of tadA* was incorporated.
Insertion of tadA-GGS	Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *tadA-GGS-T7*]	5'-drApol::kanccdB and 3'-drApol::kanccdB	tadA-Forward and tadA-GGS-Reverse used to amplify from pCMV-ABE7.10	Replacement of rApo1 in the MutaT7 strain with tadA, an adenosine deaminase.[11] Note, only a single monomer of tadA* was incorporated.

Deleting the *motAB* and *csgABCDEFG* operons through DIRex lambda *red* recombineering to decrease biofilm formation in bioreactor experiments: Deletions of the *motAB* operon[63] and the *csgABCDEFG*[64] 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.[65] 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 (**Table 4.5**). 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 (**Table 4.5**).

Separation of rApo1–T7 fusion (rApo1+T7) through DIRex lambda *red* recombineering: In order to generate a non-fusion control strain in which rApol (or drApol) and T7 are expressed separately from the same operon under the $P_{A1lacO-Tenth}$

promoter, one-step DIRex lambda red recombineering was used to insert a stop codon at the end of the rApo1 gene. The rApo1Stop targeting half-cassettes were amplified from R6K-AmilCP-kan-ccdB using the primers rApo1StopDF and AmilCP-KanR and from R6K-kan-ccdB-AmilCP using the primers rApo1StopDR and KanF-AmilCP (**Table 4.5**). The rApo1Stop half cassettes were co-electroporated to insert a stop codon after rApo1 followed by a kan-ccdB cassette flanked by large AmilCP inverted repeats nested between short 30 bp direct repeats. Excision of the AmilCP-kan-ccdB-AmilCP cassette was selected for using ccdB counterselection to obtain a markerless insertion of a stop codon after rApo1.

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 (**Table 4.3**) ($Strep^R$) 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 L-arabinose (with 10 µg/mL chloramphenicol for the MP6 strain, **Table 4.3**), 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 L-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 (**Table 4.5**). The amplicons were Sanger-sequenced using the primer 1197 (**Table 4.5**).

LacZα activity assay for quantifying T7 and MutaT7 processivity: In order to determine if the fusion of rApo1 to the N-terminus of T7 RNA polymerase affected the processivity and/or activity of the T7 RNA polymerase, the expression of the lacZα fragment from T7 promoters of varying upstream distances was measured via the cleavage of *o*-Nitrophenyl-β-galactoside (*o*NPG) using an assay adapted from a previous publication.[30] LacZα reporter plasmids C1A through C1F ($Chlor^R$) were transformed into the *ung*+, drApo1–T7 *ung*$^+$ and drApo1+T7 *ung*$^+$ strains (**Table 4.3**) and plated on LB agar with 25 µg/mL chloramphenicol and grown at 37 °C for 24 h in order to obtain clones. Colonies of each reporter/strain combination were picked in triplicate and grown in 200 µL LB with 25 µg/mL chloramphenicol and 1 mM IPTG in a parafilm-wrapped 96-well plate that was shaken at 220 r.p.m. at 30 °C for 22 h. IPTG was added to induce the expression of the lacZω fragment from the genome that complements the lacZα fragment, and to increase the expression of drApo1–T7 and T7 from the $P_{A1lacO\text{-}Tenth}$ promoter. 80 µL of each overnight culture was mixed with 120 µL Bgal mix (60 mM sodium dibasic, 40 mM sodium phosphate monobasic, 10 mM potassium chloride,1 mM magnesium sulfate, 26 mM 2-mercaptoethanol, 166 µg/mL egg-white lysozyme, 1.0 mg/mL oNPG, and 6.7% PopCulture lysis reagent) in a black, clear-bottomed 96-well plate. The OD600 and OD420 of each well was measured every 2 min over the course of 1 h in a Biotek Synergy H1 hybrid plate reader followed by double orbital shaking at 559 r.p.m. at 30 °C. The *o*NPG cleavage activity of each well was calculated by measuring the slope of the linear

region of each OD420 trace, dividing by the initial OD600 reading, and multiplying by 1000. The mean and standard deviation of each set of triplicates were calculated.

Episomal *folA* directed evolution assay to assess false positive frequency: To assess the effect that targeted versus global mutagenesis has on the false positive frequency of a directed evolution experiment, we designed a model drug resistance evolution experiment where the rate of true positive evolution corresponds to the frequency that drug resistance-conferring mutations appear in an episomal copy of a drug-sensitive gene. To create this system, the folA+T7 promoter plasmid (Amp^R)—which contains the complete, endogenous *folA* promoter and coding sequence for dihydrofolate reductase followed by a T7 promoter pointing in the reverse direction—was transformed into MutaT7 and MP6 mutagenic strains (**Table 4.3**) ($Strep^R$). These strains 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 L-arabinose (with 10 µg/mL chloramphenicol for the MP6 strain (**Table 4.3**)), 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 L-arabinose. 50 µL of a 100-fold dilution of each resuspension was plated on LB agar plates with 5 µg/mL trimethoprim (TMP) and 50 µg/mL tetrazolium chloride. 13–15 colonies from each plate were picked for colony PCR. Episomal *folA* was amplified and Sanger sequenced using the primers Alof-T7 S and 1197 (**Table 4.5**).

Bacterial growth assay measuring trimethoprim drug resistance: Isolates were grown to stationary phase following overnight incubation at 37 °C in LB with 100 µg/mL ampicillin. Cultures were diluted 1:100 into a plate containing LB broth with increasing concentrations of TMP ranging from 1 µM to 1 mM. Growth of diluted samples was determined by measuring OD600 every 5 min in a Biotek Synergy H1 hybrid plate reader followed by orbital shaking at 282 r.p.m. and incubation at 37 °C. Maximal growth rate was determined by performing "Max V" calculation in Gen5 software, using a 5-point segment of each growth curve corresponding to the highest linear slope. Upon determining maximum growth rate within each sample, growth rates were normalized to the highest growth rate within each sample series yielding the relative growth rate at each TMP concentration (**Figure 4.11**).

Table 4.3 Strains engineered and used in this work.

Strain	Genotype*
DH10B[59]	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*::[$P_{A1lacO-Tenth}$ *rApo1*]
drApo1	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *drApo1*]
MutaT7	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *MutaT7*]
drApo1–T7	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *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*::[$P_{A1lacO-Tenth}$ *MutaT7*]
ung+	DH10B *ung*⁺ Δ*motAB* Δ*csgABCDEFG*
drApo1–T7 *ung*⁺	DH10B *ung*⁺ Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *drApo1-T7*]
drApo1+T7 *ung*⁺ (Note: drApo1 and T7 expressed as separate proteins)	DH10B *ung*⁺ Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *drApo1 T7*]
ugi rApo1	DH10B *ung*⁺ Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *ugi rApo1*]
ugi MutaT7	DH10B *ung*⁺ Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *ugi MutaT7*]
ugi drApo1–T7	DH10B *ung*⁺ Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *ugi drApo1-T7*]
tadA-Only	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *tadA*]
tadA-GGS-T7	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *tadA-GGS-T7*]
tadA-XTEN-T7	DH10B Δ*ung* Δ*motAB* Δ*csgABCDEFG* [P_{lacI}<>P_{tac}] Δ(*araA-leu*)*7697*::[$P_{A1lacO-Tenth}$ *tadA-XTEN-T7*]

* 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.

Table 4.4 Mutation assay data for Figure 4.7, Figure 4.9, and Figure 4.10.

Strain	Reporter Plasmid	Resistance	CFU/mL	Standard Deviation	Frequency (Antibiotic^R CFU/mL)/(Amp^R CFU/mL)	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		
Δ*ung*	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		
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		
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	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

	+ terminator array	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 amicillin resistant cells immediately after EMS treatment.

Table 4.5 Primers and oligonucleotides used in this work.

Primer Name	Sequence
5′–Ung kanccdB	5′-gcagttaagctaggcggattgaagattcgcaggagagcgagatggctaacccctcatcagtgccaacatagta ag-3′
3′–Ung kanccdB	5′-agccgggtggcaactctgccatccggcatttccccgcaaatttactcactccgctcattaggcgggc-3′
delUng S	5′-gcagttaagctaggcggattgaagattcgcaggagagcgagatggctaacagtgagtaaatttgcggggaaatgc cggatggcagagttgccacccggct-3′
delUng AS	5′-agccgggtggcaactctgccatccggcatttccccgcaaatttactcactgttagccatctcgctctcctgcgaatcttca atccgcctagcttaactgc-3′
5′–pLacI::kanccdB	5′-cgttactggtttcacattcaccaccctgaattgactctcttccgggcgctccctcatcagtgccaacatagtaag-3′
3′–pLacI::kanccdB	5′-tggtggccggaaggcgaagcggcatgcatttacgttgacaccatcgaatgccgctcattaggcgggc-3′
pLacI::pTac AS	5′-attcaccaccctgaattgactctcttccgggcgctcattatacgagccgatgattaattgtcaacattcgatggtgtcaac gtaaatgcatgccgcttc-3′
pLacI::pTac S	5′-gaagcggcatgcatttacgttgacaccatcgaatgttgacaattaatcatcggctcgtataatgagcgcccggaagag agtcaattcagggtggtgaat-3′
delcsgDF	5′-tttcgcttaaacagtaaaatgccggatgataattccggctttttatctgtttgttgaaatatcggaattaaaaaagaattca aaaaaaagcccgc-3′
delcsgDR	5′-gcagcagaccattctctccagattcatcttatgctcgatatttcaacaaacagataaaaaagccggaattaaaaaaga attcaaaaaaaagcccgc-3′
delmotDF	5′-cttcatcaaaaaatgtctgataaaaatcgcttatatccatgctcacgctggacatcatccttccagaattaaaaaagaat tcaaaaaaaagcccgc-3′
delmotDR	5′-cgcctgacgactgaacatcctgtcatggtcaacagtggaaggatgatgtccagcgtgagcatggagaattaaaaa agaattcaaaaaaaagcccgc-3′
KanF–AmilCP	5′-gctcgacgttgtcactgaagc-3′
AmilCP–KanR	5′-cgccgtcgggcatgc-3′
5′–drApoI::kanccdB	5′-gtcgtcactctatctggcgtcacacctctcagaacaccaacaaacacgttccgctcattaggcgggc-3′
3′–drApoI::kanccdB	5′-ttcgggcagaagtaacgttcggtggtgaatttttcgatgaagttaacttccctcatcagtgccaacatagtaag-3′
drApoI S	5′-gtcgtcactctatctggcgtcacacctctcagaacaccaacaaacacgttcaagttaacttcatcgaaaaattcacca ccgaacgttacttctgcccgaa-3′
drApoI AS	5′-ttcgggcagaagtaacgttcggtggtgaatttttcgatgaagttaacttgaacgtgtttgttggtgttctgagaggtgtgac gccagatagagtgacgac-3′
dAraLeu7697 kanccdB F	5′-cagcagcagaacgcccggcacgtgctctgccagttgttcaatggcctgatccctcatcagtgccaacatagtaag-3′
dAraLeu7697 kanccdB R	5′-tgagcaggcaatcagcagttgataaccccgttgccgcgcctggcgttcaaccgctcattaggcgggc-3′
dAraLeu7697–rApoI	5′-cagcagcagaacgcccggcacgtgctctgccagttgttcaatggcctgattcgagtttatggctagctcagtcc-3′
dAraLeu7697–T7	5′-tgagcaggcaatcagcagttgataaccccgttgccgcgcctggcgttcaactgaaaatcttctctcatccgcc-3′
5′–prApoI::kanccdB	5′-gcagaacgcccggcacgtgctctgccagttgttcaatggcctgattcgagccgctcattaggcgggc-3′
3′–prApoI::kanccdB	5′-cgacgacgcagggtcgggtcaaccgcaaccggaccggtttcagaagacatccctcatcagtgccaacatagta ag-3′
PA1lacO–1 F	5′-gcagaacgcccggcac-3′
PA1lacO–1 R	5′-cgacgcagggtcggg-3′
pA1lacO1–HA Tenth gene block	5′-gcagaacgcccggcacgtgctctgccagttgttcaatggcctgattcgagaaagagtgtttatttgtgagcggataaca attacaattagattcaattgtgagcggataacaatttcacacaggctagcgaattcgagctccctctagaaataattttgttta actttaagaaggagatataccatgggcagcagctacccatacgacgtaccagattacgctatgtcttctgaaaccggtccg gttgcggttgacccgaccctgcgtcgtcg-3′
NheI–UUCG–BamHI S	5′-ctagccagcttgggtctccctaggtcgagctccgtcgacctagcataaccccgcggggcctcttcgggggtctcgcgg ggtttttttgctgaaagg-3′
NheI–UUCG–BamHI AS	5′-gatcccttcagcaaaaaaccccgcgagacccccgaagaggccccgcggggttatgctaggtcgacggagctcga cctagggagacccaagctgg-3′

1493	5′-aaaaaaaagcttgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaac cacgggatcggccattgaac-3′
1494	5′-aaaaaattaattaagctctagagaattgatcccctcag-3′
2165	5′-ttgacaattaatcatcggctcgaagcttg-3′
1197	5′-aaaaaaggatccttcgccattcaggctgcg-3′
2062	5′-aaagtgccacctggcgg-3′
Ung-Forward	5′-gtgtatatatactcctcaaacacccttgaatctttg-3′
Ung-Reverse	5′-attgacggtacgggcaacg-3′
5′–ugi kanccdB	5′-atatctccttcttaaagttaaacaaaattatttctagagggagctcgaatccctcatcagtgccaacatagtaag-3′
3′–ugi kanccdB	5′-gatacttagattcaattgtgagcggataacaatttcacacaggctagcgaccgctcattaggcgggc-3′
Ugi-Forward	5′-gatacttagattcaattgtgagcggataacaatttcacacaggctagcgattaggaggaatttcaacatgacaaatttat ctgacatc-3′
Ugi-Reverse	5′-atatctccttcttaaagttaaacaaaattatttctagagggagctcgaatcttataacattttaattttattttctccattactgtct tg-3′
rApo1StopDF	5′-gccactaccagcgtctgccgccgcacatcctgtgggcgaccggtctgaaataaggagtggcggtagcggaattaa aaaagaattcaaaaaaaagcccgc-3′
rApo1StopDR	5′-gttcatatggtatcctcttgagctcccactccctccgctaccgccactccttatttcagaccggtcgcgaattaaaaaaga attcaaaaaaaagcccgc-3′
Alof-T7 S	5′-gctcgaattctaatacgactcactataggg-3′
tadA-Forward	5′-ggagatataccatgggcagcagctacccatacgacgtaccagattacgcttccgaagtcgagttttcccatgag-3′
tadA-Only-Reverse	5′-tgagcaggcaatcagcagttgataaccccgttgccgcgcctggcgttcaagtcagtagaggattgtgctttttctgg g-3′
tadA-XTEN-Reverse	5′-tcagagaagtcgttcttagcgatgttaatcgtgttcatatggtatcctctagaaccaccagaagaaccaccagaacttt cgggtgtg-3′
tadA-GGS-Reverse	5′-cctcttgagctcccactccctccgctaccgccactccctccgctaccgccgtcagtagaggattgtgctttttctggg-3′

4.7 Acknowledgements
This work was supported by the NIH Director's New Innovator Award 1DP2GM119162 and the 56th Edward Mallinckrodt Jr. Foundation Faculty Scholar Award (both to M.D.S.). This work was also supported in part by the National Institute of Environmental Health Sciences of the NIH under award P30-ES002109. C.L.M. and L.J.P. were supported by National Science Foundation Graduate Research Fellowships under Grant No. 1122374. The authors acknowledge Vincent L. Butty (MIT) for assistance with sequencing data analysis.

4.8 References

[1] Cupples, C. G., and Miller, J. H. (1989) A set of lacZ mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions, *Proc. Natl. Acad. Sci. U.S.A. 86*, 5345–5349.

[2] Tessman, I., Ishiwa, H., and Kumar, S. (1965) Mutagenic effects of hydroxylamine in vivo, *Science 148*, 507–508.

[3] Greener, A., Callahan, M., and Jerpseth, B. (1997) An efficient random mutagenesis technique using an E. coli mutator strain, *Mol. Biotechnol. 7*, 189–195.

[4] Badran, A. H., and Liu, D. R. (2015) Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra, *Nat. Commun. 6*, 8425.

[5] Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabási, A. L., Oltvai, Z. N., and Osterman, A. L. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655, *J. Bacteriol. 185*, 5673–5684.

[6] Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., and Sabatini, D. M. (2015) Identification and characterization of essential genes in the human genome, *Science 350*, 1096–1101.

[7] Tizei, P. A., Csibra, E., Torres, L., and Pinheiro, V. B. (2016) Selection platforms for directed evolution in synthetic biology, *Biochem. Soc. Trans. 44*, 1165–1175.

[8] Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, *Nature 533*, 420–424.

[9] Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., and Kondo, A. (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems, *Science 353*, aaf8729-8721.

[10] Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., Kim, Y. B., Badran, A. H., and Liu, D. R. (2017) 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.

[11] Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage, *Nature 551*, 464–471.

[12] Kim, D., Lim, K., Kim, S. T., Yoon, S. H., Kim, K., Ryu, S. M., and Kim, J. S. (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases, *Nat. Biotechnol. 35*, 475–480.

[13] Hess, G. T., Fresard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., Montgomery, S. B., and Bassik, M. C. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells, *Nat. Methods 13*, 1036–1042.

[14] Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells, *Nat. Methods 13*, 1029–1035.

[15] Camps, M., Naukkarinen, J., Johnson, B. P., and Loeb, L. A. (2003) Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I, *Proc. Natl. Acad. Sci. U.S.A. 100*, 9727–9732.
[16] Allen, J. M., Simcha, D. M., Ericson, N. G., Alexander, D. L., Marquette, J. T., Van Biber, B. P., Troll, C. J., Karchin, R., Bielas, J. H., Loeb, L. A., and Camps, M. (2011) Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication, *Nucleic Acids Res. 39*, 7020–7033.
[17] Crook, N., Abatemarco, J., Sun, J., Wagner, J. M., Schmitz, A., and Alper, H. S. (2016) In vivo continuous evolution of genes and pathways in yeast, *Nat. Commun. 7*, 13051.
[18] DiCarlo, J. E., Conley, A. J., Penttila, M., Jantti, J., Wang, H. H., and Church, G. M. (2013) Yeast oligo-mediated genome engineering (YOGE), *ACS Synth. Biol. 2*, 741–749.
[19] Ravikumar, A., Arrieta, A., and Liu, C. C. (2014) An orthogonal DNA replication system in yeast, *Nat. Chem. Biol. 10*, 175–177.
[20] Rong, M., He, B., McAllister, W. T., and Durbin, R. K. (1998) Promoter specificity determinants of T7 RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A. 95*, 515–519.
[21] Thiel, V., Herold, J., Schelle, B., and Siddell, S. G. (2001) Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus, *J. Gen. Virol. 82*, 1273–1281.
[22] Ramiro, A. R., Stavropoulos, P., Jankovic, M., and Nussenzweig, M. C. (2003) Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand, *Nat. Immunol. 4*, 452–456.
[23] Chen, X., Zaro, J. L., and Shen, W. C. (2013) Fusion protein linkers: property, design and functionality, *Adv. Drug Deliv. Rev. 65*, 1357–1369.
[24] Camsund, D., Heidorn, T., and Lindblad, P. (2014) Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium, *J. Biol. Eng. 8*.
[25] Glascock, C. B., and Weickert, M. J. (1998) Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*, *Gene 223*, 221–231.
[26] Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H. E., Lindahl, T., and Barnes, D. E. (2000) Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication, *Mol. Cell 5*, 1059–1065.
[27] Alsoe, L., Sarno, A., Carracedo, S., Domanska, D., Dingler, F., Lirussi, L., SenGupta, T., Tekin, N. B., Jobert, L., Alexandrov, L. B., Galashevskaya, A., Rada, C., Sandve, G. K., Rognes, T., Krokan, H. E., and Nilsen, H. (2017) Uracil accumulation and mutagenesis dominated by cytosine deamination in CpG dinucleotides in mice lacking UNG and SMUG1, *Sci. Rep. 7*, 7199.
[28] Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S. (2002) AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification, *Nature 418*, 99–103.
[29] Navaratnam, N., Bhattacharya, S., Fujino, T., Patel, D., Jarmuz, A. L., and Scott, J. (1995) 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.
[30] Schaefer, J., Jovanovic, G., Kotta-Loizou, I., and Buck, M. (2016) Single-step method for beta-galactosidase assays in Escherichia coli using a 96-well microplate reader, *Anal. Biochem. 503*, 56–57.
[31] Hecht, A., Glasgow, J., Jaschke, P. R., Bawazer, L. A., Munson, M. S., Cochran, J. R., Endy, D., and Salit, M. (2017) Measurements of translation initiation from all 64 codons in E. coli, *Nucleic Acids Res. 45*, 3615-3626.
[32] Clarke, M. L., Burton, R. L., Hill, A. N., Litorja, M., Nahm, M. H., and Hwang, J. (2010) Low-cost, high-throughput, automated counting of bacterial colonies, *Cytom. A 77*, 790–797.
[33] Geissmann, Q. (2013) OpenCFU, a new free and open-source software to count cell colonies and other circular objects, *PLoS One 8*, e54072.
[34] Garibyan, L. (2003) Use of the rpoB gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome, *DNA Repair 2*, 593–608.
[35] Nilsson, A. I., Berg, O. G., Aspevall, O., Kahlmeter, G., and Andersson, D. I. (2003) Biological costs and mechanisms of fosfomycin resistance in Escherichia coli, *Antimicrob. Agents Chemother. 47*, 2850–2858.
[36] Acar, J. F., and Goldstein, F. W. (1982) Genetic aspects and epidemiologic implications of resistance to trimethoprim, *Rev. Infect. Dis. 4*, 270–275.
[37] Herrington, M. B., MacRae, T. J., Panagopoulos, D., and Wong, S.-H. (2002) A mutation in the folA promoter delays adaptation to minimal medium by *Escherichia coli* K-12, *J. Basic Microbiol. 42*, 172–180.
[38] Yin, Y. W., and Steitz, T. A. (2004) The structural mechanism of translocation and helicase activity in T7 RNA polymerase, *Cell 116*, 393–404.
[39] Schellenberger, V., Wang, C. W., Geething, N. C., Spink, B. J., Campbell, A., To, W., Scholle, M. D., Yin, Y., Yao, Y., Bogin, O., Cleland, J. L., Silverman, J., and Stemmer, W. P. (2009) A recombinant polypeptide extends the *in vivo* half-life of peptides and proteins in a tunable manner, *Nat. Biotechnol. 27*, 1186–1190.
[40] Mairhofer, J., Wittwer, A., Cserjan-Puschmann, M., and Striedner, G. (2015) Preventing T7 RNA polymerase read-through transcription-A synthetic termination signal capable of improving bioprocess stability, *ACS Synth. Biol. 4*, 265–273.
[41] Serrano-Heras, G., Ruiz-Maso, J. A., del Solar, G., Espinosa, M., Bravo, A., and Salas, M. (2007) Protein p56 from the *Bacillus subtilis* phage phi29 inhibits DNA-binding ability of uracil-DNA glycosylase, *Nucleic Acids Res. 35*, 5393–5401.
[42] Thuronyi, B. W., Koblan, L. W., Levy, J. M., Yeh, W. H., Zheng, C., Newby, G. A., Wilson, C., Bhaumik, M., Shubina-Oleinik, O., Holt, J. R., and Liu, D. R. (2019) Continuous evolution of base editors with expanded target compatibility and improved activity, *Nat. Biotechnol. 37*, 1070–1079.
[43] Abdullah, T., Faiza, M., Pant, P., Rayyan Akhtar, M., and Pant, P. (2016) An analysis of single nucleotide substitution in genetic codons - probabilities and outcomes, *Bioinformation 12*, 98–104.
[44] Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017) Dynamic RNA modifications in gene expression regulation, *Cell 169*, 1187–1200.

[45] Wang, Z., and Mosbaugh, D. W. (1989) Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase, *J. Biol. Chem. 264*, 1163–1171.
[46] Meyer-Ficca, M. L., Meyer, R. G., Kaiser, H., Brack, A. R., Kandolf, R., and Kupper, J. H. (2004) Comparative analysis of inducible expression systems in transient transfection studies, *Anal. Biochem. 334*, 9–19.
[47] Conrad, B., Savchenko, R. S., Breves, R., and Hofemeister, J. (1996) *A T7 promoter-specific, inducible protein expression system for* Bacillus subtilis, *Mol. Gen. Genet. 250*, 230–236.
[48] McBride, K. E., Schaaf, D. J., Daley, M., and Stalker, D. M. (1994) 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.
[49] Lieber, A., Sandig, V., and Strauss, M. (1993) A mutant T7 phage promoter is specifically transcribed by T7-RNA polymerase in mammalian cells, *Eur. J. Biochem. 217*, 387–394.
[50] Weinstock, M. T., Hesek, E. D., Wilson, C. M., and Gibson, D. G. (2016) *Vibrio natriegens* as a fast-growing host for molecular biology, *Nat. Methods 13*, 849–851.
[51] Dower, K., and Rosbash, M. (2002) T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export, *RNA 8*, 686–697.
[52] Moore, C. L., Papa, L. J., III, and Shoulders, M. D. (2018) A processive protein chimera introduces mutations across defined DNA regions *in vivo*, *J. Am. Chem. Soc. 140*, 11560–11564.
[53] Wycuff, D. R., and Matthews, K. S. (2000) Generation of an AraC-araBAD promoter-regulated T7 expression system, *Anal. Biochem. 277*, 67–73.
[54] Miller, A. W., Befort, C., Kerr, E. O., and Dunham, M. J. (2013) Design and use of multiplexed chemostat arrays, *J. Vis. Exp.*, e50262.
[55] Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM., *arXiv preprint arXiv:1303.3997*.
[56] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools, *Bioinformatics 25*, 2078–2079.
[57] Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics 27*, 2987–2993.
[58] Wang, H., Bian, X., Xia, L., Ding, X., Muller, R., Zhang, Y., Fu, J., and Stewart, A. F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection, *Nucleic Acids Res. 42*, e37.
[59] Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., Ayele, M., Gibbs, R. A., Csorgo, B., Posfai, G., Weinstock, G. M., and Blattner, F. R. (2008) The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse, *J. Bacteriol. 190*, 2597–2606.

[60] Duncan, B. K. (1985) Isolation of insertion, deletion, and nonsense mutations of the uracil-DNA glycosylase (*ung*) gene of *Escherichia coli* K-12, *J. Bacteriol. 164*, 689–695.
[61] Anderson, J. C. Anderson promoter collection. Available: http://parts.igem.org/Promoters/Catalog/Anderson.
[62] Camsund, D., Heidorn, T., and Lindblad, P. (2014) Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium, *J. Biol. Eng. 8*, 4.
[63] Pratt, L. A., and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili, *Mol. Microbiol. 30*, 285–293.
[64] Prigent-Combaret, C., Prensier, G., Le Thi, T. T., Vidal, O., Lejeune, P., and Dorel, C. (2000) Developmental pathway for biofilm formation in curli-producing *Escherichia coli* strains: role of flagella, curli and colanic acid, *Environ. Microbiol. 2*, 450–464.
[65] Näsvall, J. (2017) Direct and Inverted Repeat stimulated excision (DIRex): Simple, single-step, and scar-free mutagenesis of bacterial genes, *PLoS One 12*, e0184126.

Chapter 5: Towards a Generalizable Directed Evolution System Based on Lambda Phage Lysogeny

5.1 Author Contributions

Louis J. Papa planned research, designed experiments, and analyzed data. Matthew D. Shoulders acquired funding and supervised experiments and data analysis.

5.2 Introduction

Mimicking natural evolution through directed evolution workflows is the premier method for studying evolutionary processes and generating novel biomolecules to catalyze reactions, form intricate structures, set up sophisticated regulatory networks, and treat human diseases. Recent advances have dramatically reduced the time required for a single directed evolution cycle by integrating the mutagenesis, selection, and amplification steps into one self-sustaining process via a constantly-mutating M13 phage whose survival is linked to the activity of a biomolecule of interest (BOI) encoded by the phage in a platform known as PACE.[1]

PACE has enabled the evolution of an impressive array of biomolecules and has made a major impact in the field.[1-6] However, there are limitations of PACE that render the technique difficult to use for non-experts and constrain the nature of the biomolecules that can currently be evolved:

Limitation 1: The entire phage genome accumulates mutations, which can lead to false positives and reduce the accessible mutation rate owing to mutational error catastrophe.

Limitation 2: M13 phage can accommodate only ~2 kb of non-phage DNA before they struggle to replicate,[7] precluding the evolution of large proteins, multi-component genetic circuitry, or assembly line enzymes that produce important natural products.

Limitation 3: Building successful selection couples and maintaining appropriate selection pressure can be challenging owing to an inability to robustly and cheaply monitor selection pressure and progress in real time.

Despite the limitations of PACE, the productivity made possible by the method highlights the transformative potential accessible to any next-generation method that removes these limitations and enables the broad application of continuous directed evolution across many targets and many labs. We have begun efforts to develop such a method based on re-engineering the lambda bacteriophage as a vehicle for directed evolution rather than the M13 bacteriophage.

In the lambda-based scheme (**Figure 5.1**), plasmids that contain the necessary *cos* recognition sites (cosmids), but lack the rest of the lambda phage genome, are packaged into lambda capsids. Cosmids are packaged with the assistance of an unpackageable, *cos*-less helper prophage that is integrated into the *Escherichia coli* genome.[8] The helper lambda prophage remains dormant in the *E. coli* in a state known as "lysogenic phase," only packaging cosmids when it is induced to switch to the "lytic phase." The lambda phage cI repressor protein prevents the transcription of all phage genes except itself and thus maintains this dormant lysogenic phase.[9] Until recently, this approach was not viable owing to recombination between the helper prophage and cryptic defective prophages that occur naturally in the K-12 *E. coli* genome. Two of these cryptic prophages share significant homology with the helper lambda prophage and provide *cos* sites during recombination, thus reintroducing the *cos* site into lambda and producing replication-competent phage contamination. Such contamination dominates the culture and previously prevented any attempts at directed evolution. Fortunately, this problem was recently solved by the deletion of all nine cryptic prophages from the K-12 *E. coli* genome.[8] The main motivation for developing this specialized strain was to generate contamination-free packaged cosmids that deliver the T7 RNA polymerase for sudden off-to-on expression of toxic genes from a T7 promoter.[8] However, we adapted this

system in an effort to create a next-generation continuous directed evolution platform.

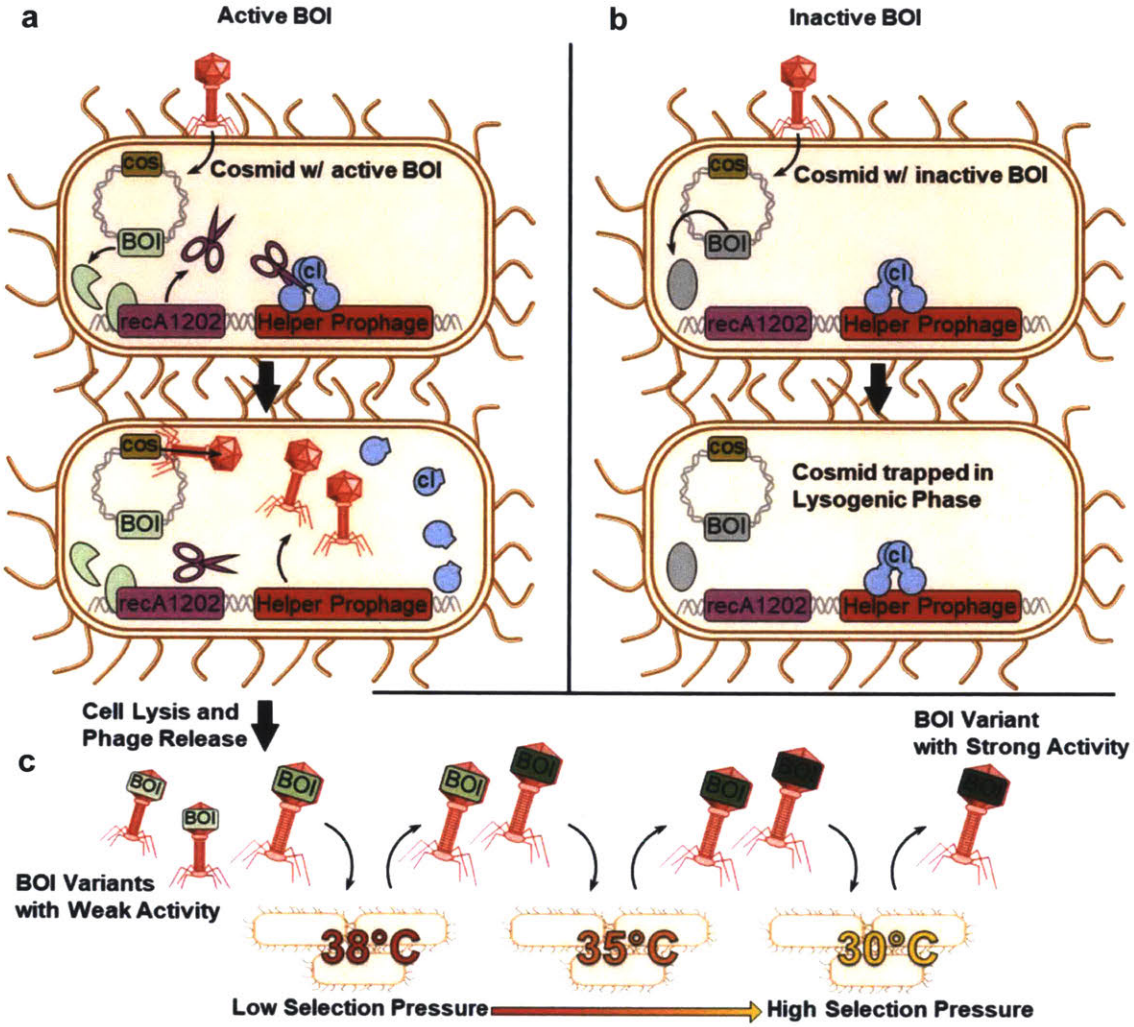

Figure 5.1 Positive selection scheme.
a) Phage particles carrying cosmids that encode active BOIs infect cells, which subsequently produce more phage particles that carry active BOI genes. Note: the helper prophage genome excises itself from the host genome upon induction of the lytic phase but, for simplicity, the excision is not depicted. **b)** Phage particles carrying cosmids encoding inactive BOIs infect cells, but these cells remain in the lysogenic phase and trap the cosmid, thus removing it from the phage population. **c)** Lowering the temperature increases selection pressure and permits only the phages carrying the most active BOI variants to survive.

In order to take advantage of this new development and harness lambda phage as an engine for continuous evolution, we rewired the natural lysogenic/lytic phase switch circuitry of lambda as a tunable selection circuit. To positively select for new desired activities, we coupled BOI activity to the expression of recA1202,[10] a hyperactive variant of the protein that causes cI cleavage and, consequently, induction of the lytic phase (**Figure 5.1a**). Note that, for simplicity, we refer to and depict recA as "cleaving" cI throughout this chapter, but recA is not actually a protease. RecA binding to cI causes cI autocleavage, meaning recA is a co-protease. Cosmids carrying inactive BOI variants

become trapped in the *E. coli* cell and fail to propagate because they cannot induce recA1202 expression (**Figure 5.1b**). In contrast, those cosmids carrying BOI variants that can induce recA1202 expression are packaged into lambda capsids and proceed to propagate from cell-to-cell (**Figure 5.1a**). The BOI gene within the cosmid continuously mutates due to the expression of mutagenic genes from the inducible MP6 plasmid.[11] On the other hand, the helper prophage is constantly replaced with its unmutated form along with its *E. coli* host. This strategy prevents evolution of the lambda phage genome itself and resolves **Limitation 1**. Additionally, because all of the native lambda phage DNA except the *cos* site is no longer carried within the phage capsid, the full 53 kb capacity of the phage capsid can be dedicated to evolving cargo.[12] Thus, the size limit of evolutionary targets in the cosmid-based system is more than an order of magnitude larger than that of M13 phage system, thus addressing **Limitation 2**.

Importantly, our data suggest that we can easily tune the stringency of selection pressure with temperature by using the temperature-sensitive variant of cI, cI857.[13] Dosable loss of repressor activity with increasing temperature lowers the threshold for induction of the lytic phase.[13] The resulting temperature-tunable selection pressure (**Figure 5.1c**) addresses **Limitation 3** by eliminating the need to construct different selection strains and optimize selection conditions through plasmid copy numbers, promoter strengths, flow rates, or chemical modulators. The use of lambda phage further addresses **Limitation 3** because, unlike M13 phage, lambda phage kills and lyses its host, thus shifting the culture from cloudy to clear and allowing for simple, cheap, real-time optical monitoring of evolutionary progress.

To negatively select against undesired activities, undesired BOI activity is linked to the expression of an uncleavable cI variant, cIG112E.[14] Cosmids carrying BOI variants with an undesired activity induce uncleavable cI expression, which traps cosmids in the lysogenic phase even if the BOI variant simultaneously induces recA1202 expression (**Figure 5.2**). Because the cleavable cIG112E variant will also be temperature-sensitive, negative selection pressure can also be tuned by temperature.

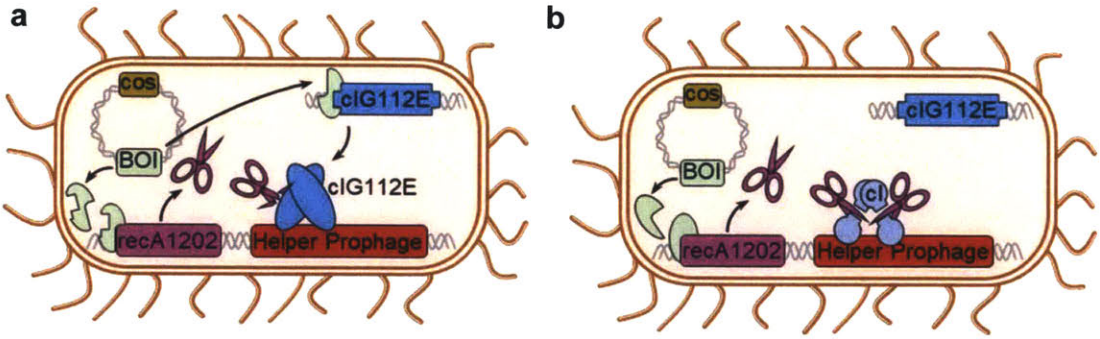

Figure 5.2 Negative selection scheme.
a) A BOI with both the desired activity and the undesired activity cannot induce the helper phage because cIG112E cannot be cleaved by recA1202. **b**) An active BOI without the undesired activity is able to induce the helper phage because cIG112E is not expressed.

While we are still integrating all the parts of the system, we expect that this cosmid-based circuit to respond to three simultaneous inputs: temperature (fully controlled by the user), recA1202 expression level (determined by the desired BOI activity), and cIG112E

expression (determined by undesired BOI activity). The output of the circuit is the number of phage particles produced (**Figure 5.3**). The tunability sets the stage for simple, robust continuous directed evolution experiments.

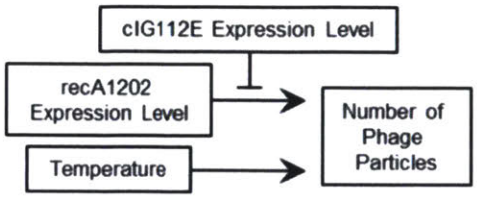

Figure 5.3 Selection circuit overview.
Higher temperature and higher recA1202 expression increase the number of phage particles, while cIG112E expression negates the effect of recA1202 expression.

Initial characterization of the positive and negative selection circuit using a plasmid lacZ expression assay was very promising. Translating the circuit to a cosmid and lysogen based system requires further optimization because the current helper prophage we are using is too stable and is not easily induced by recA1202. Fortunately, there are many variants of the lambda prophage reported in the literature with a wide range of stabilies,[9,15] so we are confident that we will be able to tune the inducibility of the prophage to the necessary level. After system optimization, we will begin experiments to evolve new variants of the T7 RNA polymerase and epothilone synthetase. If successful, this system greatly simplifies the process of setting up and tuning selection conditions, and expands the range of targets accessible to *in vivo* continuous directed evolution.

5.3 Results

5.3.1 Proof-of-principle circuitry data

As a preliminary validation of our selection circuit, we created a simplified plasmid-based, phage-less version of the system (**Figure 5.4a**). Different versions of the cI repressor were constitutively expressed at a constant level, and different versions of recA were expressed at varying levels from an arabinose inducible promoter, P_{BAD}, to simulate variable BOI activity. We measured the level of repression using a *lacZα* reporter gene under the control of a cI-repressed P_L promoter from lambda phage. We expected that increased expression levels of recA would cleave the cI repressor, leading to increased expression of lacZα (**Figure 5.4b**). LacZ activity was measured on a plate reader using a chromogenic *o*NPG cleavage assay (**Figure 5.4c**).

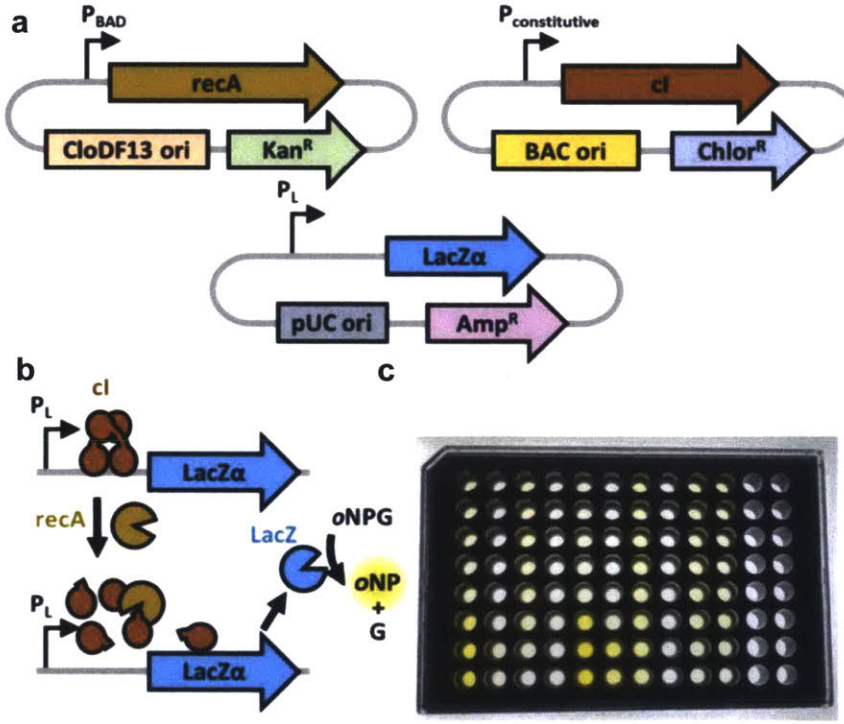

Figure 5.4 Assay for cleavage of cI by recA.
a) Maps of the three plasmids used in the assay: one to inducibly express recA variants, one to constitutively express cI variants, and one to measure cI repression via lacZα expression. **b**) RecA-mediate cleavage of cI results in lacZα expression. **c**) LacZ activity was measured on a plate reader by measuring the absorbance at 420 nm. The product of *o*NPG cleavage, *o*-nitrophenol, has a bright yellow color.

We first tested the activity of wild-type recA, hyperactive recA1202, and inactive recA1 (as a negative control) with varying levels of arabinose induction. We found that high levels of arabinose induction of hyperactive recA1202 led to high levels of lacZ activity as intended, especially when the cI857 repressor was used (**Figure 5.5**). Wild-type recA and the cI857 repressor showed some dose response between arabinose and

lacZ activity, but the wild-type cI repressor seemed resistant to derepression. Induction of the inactive recA1 variant did not significantly affect lacZ activity, as expected.

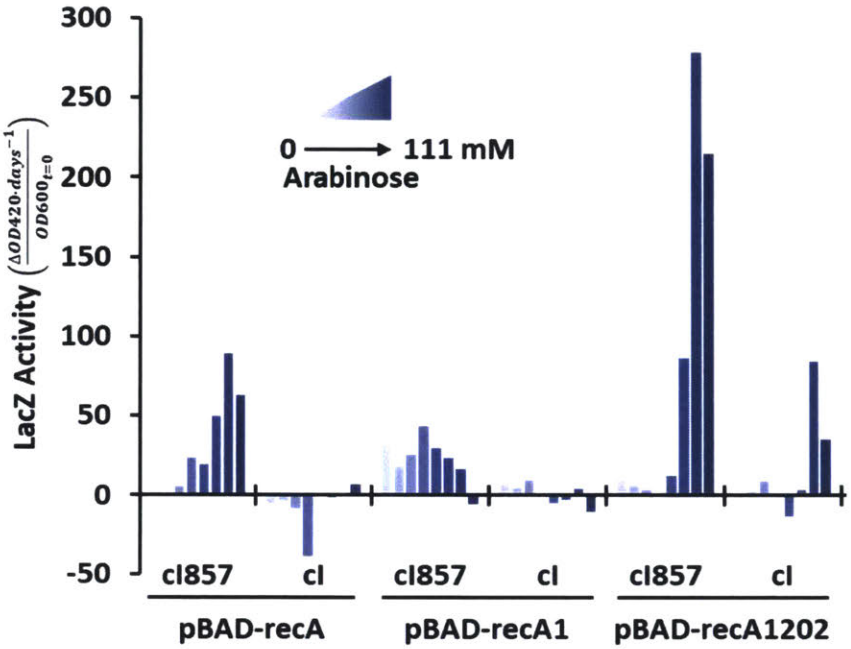

Figure 5.5 P_L phage promoter activation by recA1202.
Activity of a P_L promoter repressed by either the wild-type cI repressor or the temperature-sensitive cI857 repressor was measured using a lacZ activity assay at 32 °C. Varying expression levels of the wild-type recA, inactive recA, or hyperactive recA1202 co-proteases were induced in mid-log cultures with arabinose for 3 hours before measuring lacZ activity.

After demonstrating that recA1202 expression can dosably derepress the phage P_L promoter, we next validated that increasing the temperature can also dosably derepress P_L when using the temperature-sensitive cI857 repressor. While both the cI and cI857 repressors appeared to derepress with higher temperatures, the cI857 promoter derepressed to a greater extent and at lower temperatures than cI (**Figure 5.6**). This trend was observed in each of the three strains carrying the recA, recA1, and recA1202 plasmids in the absence of arabinose induction.

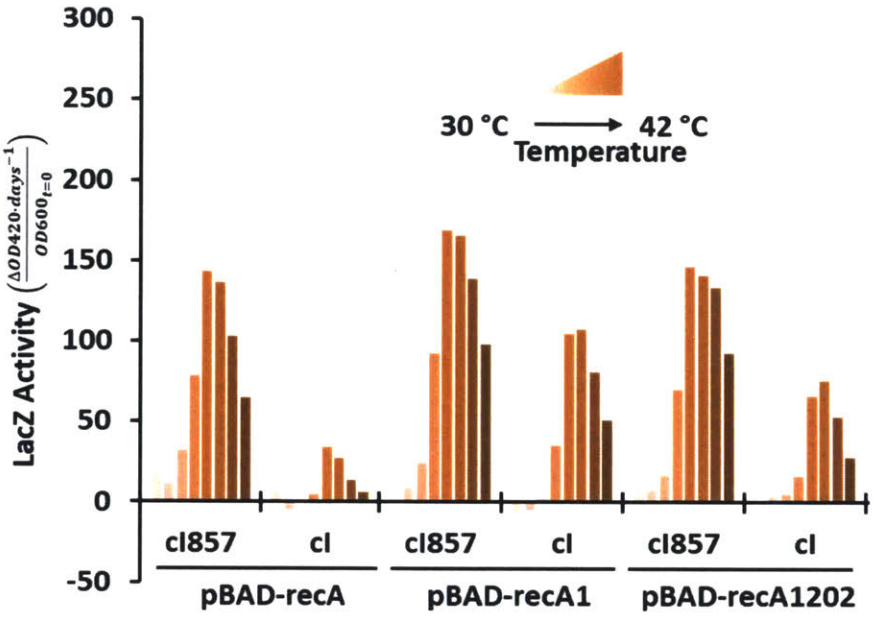

Figure 5.6 Increased temperature dosably derepresses P_L.
Activity of a P_L promoter repressed by either the wild-type cI repressor or the temperature-sensitive cI857 repressor was measured using a lacZ activity assay at temperatures ranging from 30 °C to 42 °C. LacZ activity was measured in strains carrying the wild-type recA, inactive recA, or hyperactive recA1202 expression plasmids, but without arabinose induction. Mid-log cultures were incubated at the indicated temperatures for 4 h before measuring lacZ activity.

To ensure that the effects of temperature and recA induction are additive in the circuit, we measured cI857 derepression at varying recA1202 expression levels and at varying temperatures. Indeed, the effects appeared additive over an 8 °C temperature range and two orders of magnitude of arabinose concentration (**Figure 5.7**). These preliminary results with a simplified circuit are promising and suggest that the lambda phage circuitry can be rewired to allow tunable positive selection using the temperature-sensitive cI857 repressor and the hyperactive recA1202 co-protease.

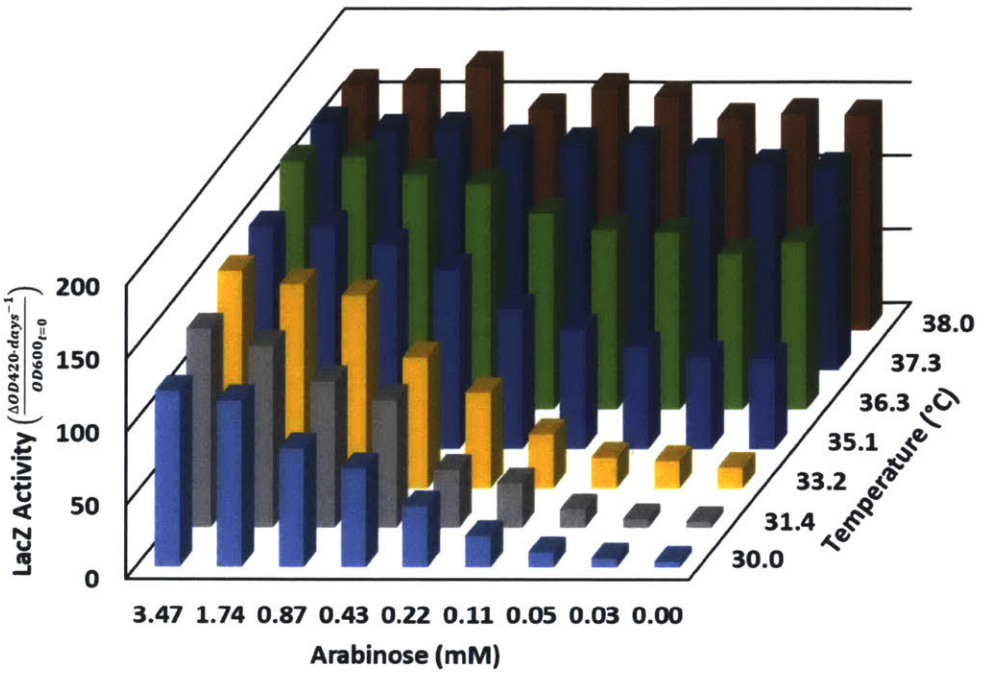

Figure 5.7 RecA1202 expression and temperature additively activate P_L. Activity of a P_L promoter repressed by a temperature-sensitive cI857 repressor. Varying expression levels of hyperactive recA1202 co-protease levels were induced in mid-log cultures with arabinose at various temperatures for 4 hours before measuring lacZ activity.

In order to test the feasibility of tunable negative selection, we tested a cI variant that should be both temperature-sensitive, and uncleavable, cI857-G112E.[14] LacZ activity was fully repressed by cI857-G112E even at high recA and recA1202 levels (**Figure 5.8a**). However, cI857-G112E was still responsive to temperature, suggesting that the strength of negative selection can still be tuned (**Figure 5.8b**).

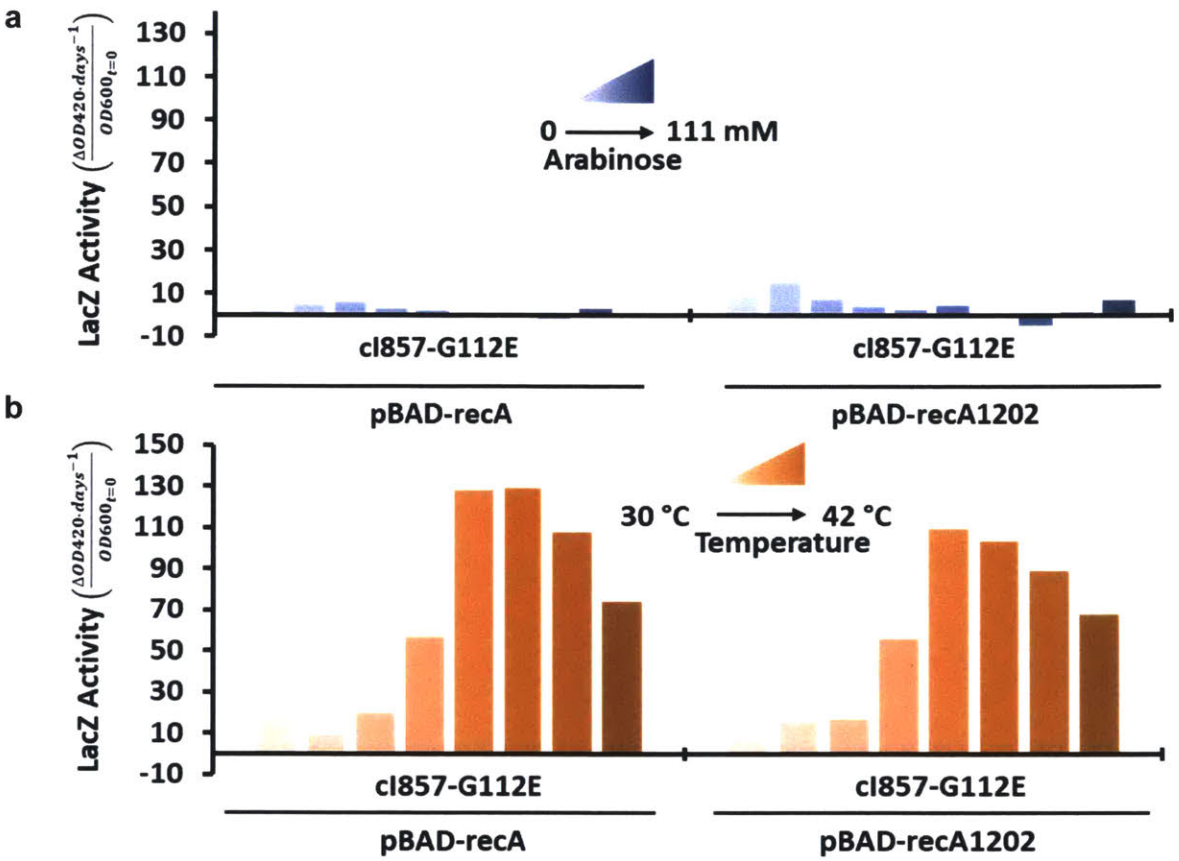

Figure 5.8 The cI857-G112E repressor is resistant to recA1202, but temperature-sensitive.

(**a**) Activity of a P_L promoter repressed by the cI857-G112E repressor at 32 °C. Varying expression levels of the wild-type recA or hyperactive recA1202 co-proteases were induced in mid-log cultures with arabinose for 3 h before measuring lacZ activity.

(**b**) Activity of a P_L promoter the cI857-G112E repressor at temperatures ranging from 30 °C to 42 °C. LacZ activity was measured in strains carrying the wild-type recA or hyperactive recA1202 expression plasmids, but without arabinose induction. Mid-log cultures were incubated at the indicated temperatures for 4 h before measuring lacZ activity.

5.3.2 Selection lysogen construction

With promising initial data in hand, we next sought to recapitulate the simplified plasmid-based selection circuit with cosmids and the *cos*-less helper. To achieve this objective, we began by engineering the previously reported CY2115 *E. coli* strain.[8] Importantly, aside from harboring a *cos*-less lambda prophage and lacking all nine cryptic prophages (**Table 5.1**), the CY2115 strain is a monolysogen, meaning it only contains one copy of the prophage genome. When a lambda phage infects a cell and enters the lysogenic phase by integrating the phage genome into the host genome, it inserts multiple tandem copies of the phage genome in about 50% of all lysogens.[8,16] The *E. coli* strains harboring more than one copy of the lambda prophage are known as a polylysogens. Attempting to edit the lambda prophage genomes of a polylysogen via recombineering results in a mixture of edited and unedited genomes.[8] Furthermore, polylysogens can potentially hamper two-step recombineering via homologous recombination between *kan*^{R}+*ccdB* intermediates and unedited genomes. Polylysogeny creates a scenario that is analogous to the BAC multimerization problem discussed in Chapter 2.

Table 5.1 Genotype of the LseI selection lysogen and parental *E. coli* strains.

Strain Name	Strain Genotype	Prophage Genotype	Source
BW25113	*Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ⁻ rph-1 Δ(rhaD-rhaB)568 hsdR514*	N/A	Previously reported.[17,18]
BW25113 Δ9	BW25113 *Δrac ΔCP4-57 ΔCPS-53 ΔDLP12 ΔQin Δe14 ΔCP4-6 ΔCPZ-55 ΔCP4-44 ΔKm*	N/A	Previously reported.[19]
CY2115	BW25113 Δ9	*cI857 ind-1 Sam7 stf⁻ Δcos::FRT-cml-FRT*	Previously reported.[8]
LseI	BW25113 Δ9 *rpsL(Str^R) ΔrecA*	*cI857 S⁺ stf⁺ Δcos::FRT-cml-FRT*	This work.

We were unable to use the CY2115 strain directly because it harbored a lambda prophage with (1) the *Sam7* mutation, a premature amber stop codon that disrupts the gene encoding holin S, which is responsible for efficient cell lysis without artificial chloroform lysis,[8] and (2) the *ind-1* point mutation in the *cI* gene, which renders the cI repressor uncleavable by recA.[20] Reversing these two mutations was absolutely essential for applying our envisioned selection circuit based on cI cleavage and release of phage capsids carrying functional BOI variants. We were able to reverse the *Sam7* mutation using two step recombineering as described in Chapter 2. Eliminating the *ind-1* mutation proved more challenging because the *cI* gene is essential for repressing the lambda phage genome and preventing cell death. Thus, we could not form a *kan*^{R}+*ccdB* intermediate that interrupts the *cI857 ind-1* mutant and replace it with *cI857* in the second step.

In order to alter the essential *cI* gene, we adapted a plasmid cointegration strategy[21,22] such that the *cI* gene is never fully deleted (**Figure 5.9**). A plasmid with the desired *cI857* allele and a flanking *kan*^{R}+*ccdB* cassette was constructed with the R6K origin, which is replication-incompetent in most *E. coli* strains.[23] The R6K origin can only replicate in cells expressing the pi protein,[24] such as Pir2 *E. coli* cells, and thus can only confer antibiotic resistance to BW25113 derivatives by integrating into the *E. coli* genome. A single crossover event results in a duplication of the *cI* gene, with one parent copy and

one modified copy. We selected for such cointegrants on kanamycin plates while simultaneously expressing the ccdA antitoxin, which renders ccdB non-lethal.[25] Intramolecular recombination between the two copies of the *cI* gene excises the *kan*^{R}+*ccdB* cassette and the rest of the vector backbone, leaving behind one copy of the *cI* gene. We selected for the resolution of cointegrants into single copies of the *cI* gene by no longer expressing ccdA and thus rendering unexcised *ccdB* toxic. Selection for resolved cointegrants produced a mixture of *cI857 ind-1* and *cI857* recombinants. We screened for the latter via Sanger sequencing.

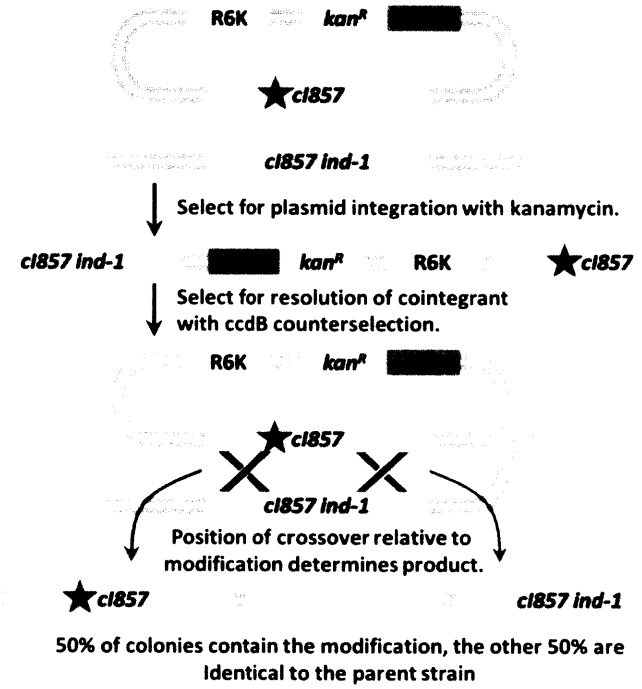

Figure 5.9 Reversing the *ind-1* mutation via plasmid cointegration.
The desired point mutation is indicated as a red star. Both the parent allele and modified allele are present after integration of the non-replicating R6K plasmid. Intramolecular recombination resolves the two copies of cI857 into one, with the position of the crossover determining whether the parent or modified base is kept.

We also made additional modifications to the CY2115 lysogen using two-step recombineering to ensure optimal performance of the platform. We deleted the endogenous *recA* gene to prevent any background cleavage of *cI857*. Thus, only BOI-induced expression of recA1202 from a plasmid can induce the helper prophage via cI cleavage. We also made the lysogen streptomycin resistant by mutating the *rpsL* gene, thus enabling the use of *rpsL* counterselection for future engineering,[26] mutation rate assays,[27] or plasmid curing. Finally, we repaired the gene encoding the side tail fiber protein, which is inactivated by a frameshift mutation in most laboratory strains of lambda phage.[28,29] Repair of the *stf* gene has been shown to increase the phage capsid adsorption rate by 8-fold.[30] We referred to the final modified lysogen that was compatible with our selection circuit as "LseI".

We note that when constructing the above series of modifications, we often found that *kan*^{R}+*ccdB* intermediates would revert to the parent sequence when attempting the

second recombineering step, similar to the BAC multimerization phenomenon described in Chapter 2. In these instances, curing and reintroducing the psc101-gbaA recombineering plasmid seemed to resolve the issue and allow us to make the modification. The psc101-gbaA plasmid shares significant homology with the lambda prophage genome, because both contain the lambda *red* recombination genes. It is likely that the lambda prophage frequently duplicated via recombination with or integration into the psc101-gbA plasmid, frustrating recombineering efforts. We also note that before embarking on the above series of modifications, we deleted the attR site of the lambda prophage genome because it is required for prophage excision.[31] We envisioned that deleting attR would prevent prophage excision and reinsertion, and thus potential tandem multimerization. It is unclear if preemptive attR deletion was necessary, and it does not appear to have prevented recombination with the psc101-gbaA plasmid. After all modifications were made, the attR deletion was repaired.

5.3.3 Initial characterization of selection lysogen LseI with the recA1202 cosmid.

With the LseI strain in hand, we next constructed a cosmid that contains the high copy pUC origin, the low copy BAC origin, the *cos* packaging signal, the ampicillin resistance gene, and recA1202 under the control of a constitutive *lac* promoter (**Figure 5.10**). The high copy pUC origin allows for high titers of packaged cosmids upon cell lysis and overrides the low copy BAC origin. The pUC origin can be excised to convert the cosmid to a BAC, which can be easily edited using recombineering. The ~4.3 kb BAC origin also serves as "stuffer" DNA to ensure that the cosmid is of sufficient size, since packing efficiency correlates strongly with cosmid size.[32] Additionally, we used the 223 bp minimal *cos* site that is required for packaging.[33] The minimal *cos* site only shares 16 bp of homology with the *cos*-less helper prophage on one side, thus preventing double-crossover homologous recombination that can generate replication-competent phage. The formation of replication-competent lambda phage has not been observed using the minimal *cos* site and CY2115.[8]

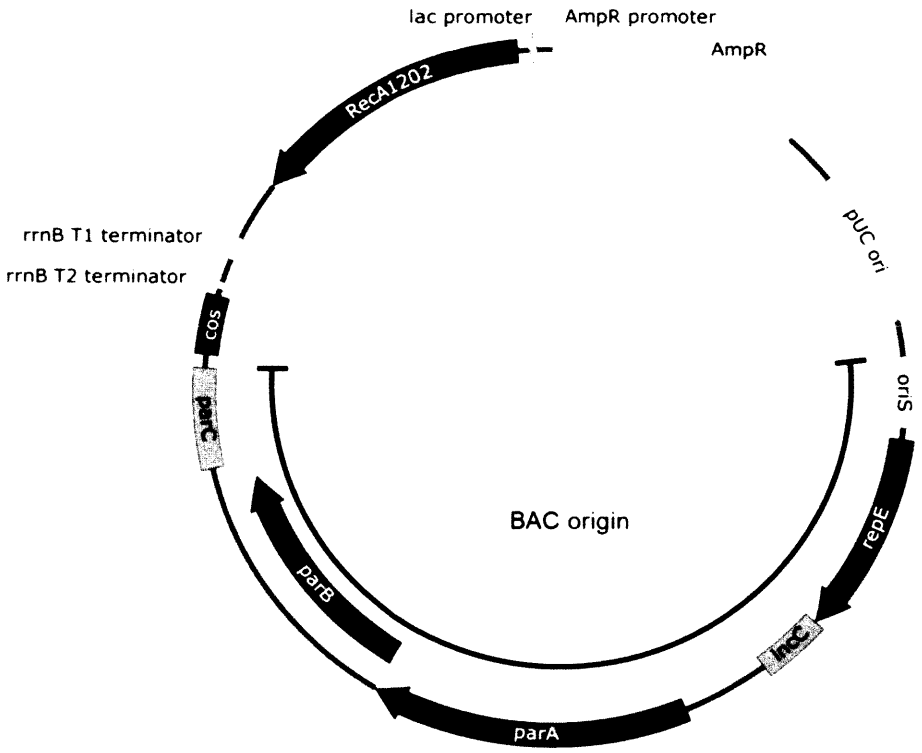

Figure 5.10 Plasmid map of the recA1202 cosmid.
This cosmid constitutively expresses recA1202 and contains a *cos* site (red) for packaging into lambda phage capsids.

Constitutive recA1202 expression from the recA1202 cosmid served as the most direct link between cosmid-borne BOI function and cosmid propagation for testing the positive selection circuit. The recA1202 cosmid was transformed into the CY2115 strain and the LseI strain. We expected that only the CY2115 could form colonies harboring the recA1202 cosmid, since recA1202 should not be able to cleave the cI857 ind-1 repressor in CY2115 but should cleave the cI857 represor in LseI and thus kill the LseI cells. We were, therefore, surprised to find that LseI could harbor the recA1202 cosmid at 30 °C

without inducing the helper prophage. The LseI strain harboring the recA1202 cosmid was induced at 42 °C to make an initial stock of packaged recA1202 cosmid. Upon infecting cosmid-less LseI cells with packaged recA1202 cosmid, we needed to raise the incubator temperature to 40 °C before observing any plaques (**Figure 5.11**). Plates incubated at 37 °C or 39 °C formed lawns with no observable plaques. While the formation of plaques shows that the LseI strain can propagate recA1202 cosmids as intended, the temperature window for observing plaques was surprisingly small. The cosmid-less LseI strain itself no longer grew at 41 °C, presumably due to lethal temperature induction of the phage without recA1202 assistance. Packaged recA1202 cosmids would likely only have a selective advantage over non-recA1202 cosmids at temperatures above 39 °C, but below 41 °C, where helper prophages would be universally induced. We found it difficult to control the temperature within such a narrow range using standard laboratory incubators.

Figure 5.11 Packaged recA1202 cosmids formed plaques on the LseI lysogen.
An LseI lysogen culture was infected with several hundred packaged recA1202 cosmids, mixed with molten top agar, and plated on an LB agar plate. The plate was incubated at 40 °C overnight until a lawn with plaques (clear spots) had formed.

We suspect that the small window of inducibility, even with high constitutive recA1202 expression, was due to the autoregulatory control of cI expression from the helper prophage, which differs from the constitutive expression of cI in the plasmid assay. Along with repressing the P_L and P_R promoters which are the main drivers of the lytic phase, cI also represses the P_{RM} promoter that drives its own expression during the lysogenic phase. When cI levels begin to drop during lysogeny, the P_{RM} promoter becomes derepressed before P_L or P_R, and expresses cI until levels are sufficient to suppress P_{RM} once again.[34] This feedback loop is likely the reason the prophage seems resistant to recA1202 based induction. We hypothesize that recA1202 and cI establish an equilibrium at lower temperatures that does not surpass the threshold for induction, with recA1202 cleavage not able to keep up with expression of new cI molecules. However, when the temperature is raised we suspect this lowers the functional levels of

the temperature-sensitive cI857 variant and tips the balance in favor of recA1202 cleavage and induction.

Ideally, we want the helper prophage to be easily inducible at low temperatures when recA1202 levels are high. We also want low levels of recA1202 expression to induce the prophage with the assistance of slightly higher temperature. Currently, we require high recA1202 levels and high temperature to induce the prophage, suggesting that the barrier to prophage induction is too high to provide dosable selection—although selection is certainly possible. Fortunately, decades of research with lambda phage has resulted in the characterization of countless lambda phage mutants and a deep knowledge of lambda phage mechanics. There are numerous mutations reported in the literature that make the lysogenic state more unstable by weakening the P_R or P_{RM} promoter or by mutating the cI repressor.[9,15,35] We are currently constructing new helper lysogens and will test the stability of the lysogenic state in each.

5.4 Discussion

Initial results are promising, showing that we can couple the function of a cosmid-borne BOI to induction of a helper lambda prophage and subsequent cosmid packaging. However, we still need to optimize the genotype of the helper prophage, because the current helper prophage is too stable and difficult to induce. After tuning the stability of the prophage with previously reported mutations,[9,15] we will test the dynamic range of selection by expressing recA1202 at various levels and at various temperatures, by quantifying the titer of packaged cosmids produced. We expect that recA1202 expression and temperature will additively and dosably increase the titer of packaged cosmids, recapitulating the results seen with the lacZ expression assay. We will also confirm that expression of the uncleavable cIG112E variant overrides recA1202 expression, but not temperature induction.

After validating the dosability of the optimized selection circuit, we will pursue an initial proof-of-principle experiment in which we will evolve the T7 RNA polymerase to recognize the T3 promoter rather than the T7 promoter. This experiment was previously carried out using PACE,[3] so we know that this particular evolution is possible. By placing the recA1202 gene downstream of a T3 promoter, we will positively select for RNA polymerase variants that recognize the T3 promoter and thus induce the lytic phase. We will begin at a high temperature to allow variants with weak activity for the T3 promoter to propagate, as in **Figure 5.1c**. As we lower the temperature, selection stringency will increase, selecting for variants with increasing activity for the T3 promoter. Once variants with strong activity for the T3 promoter are obtained, as evidenced by cosmid propagation at low temperatures, we will negatively select against any residual activity for the T7 promoter by placing the uncleavable cIG112E gene downstream of a T7 promoter. Under this negative selection scheme, variants must still induce recA1202 expression from the T3 promoter while simultaneously not inducing cIG112E expression from the T7 promoter. As with positive selection, negative selection pressure will be modulated by starting at a high temperature and then gradually increasing negative selection stringency by lowering the temperature. Successful execution of this experiment would demonstrate that all components of the system are working properly and pave the way for more challenging evolutions with biotechnological relevance.

We are particularly interested in the evolution of non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs). NRPSs and PKSs can produce extremely complex, synthetically challenging molecules, many of which are life-saving antibiotics or chemotherapeutics used in the clinic today. Researchers have tried for decades to engineer and evolve assembly line NRPS/PKS enzymes to produce complex novel compounds by rearranging modules to generate chimeric assembly lines, but most of these efforts fail or generate enzymes with minimal activity. A major challenge is that module–module interactions or linkages are critical to the function of the enzymes, and individual modules taken out of their usual context often function poorly.[36,37] Directed evolution could play a role in adapting rearranged modules to their new context, but NRPS/PKS assembly line complexes are generally very large, complicating their evolution by plasmid transformation or by M13 phage-based continuous evolution. Large plasmids greater than 10 kb transform poorly by heat-shock transformation[38] or electroporation,[39] and the M13 bacteriophage replicates poorly with more than 2 kb of non-phage DNA inserted into its genome.[7] If successful, our lambda-based system with

its large BOI capacity would be uniquely well-suited for evolving large NRPSs and PKSs.

As one of the exciting applications of this system, we will attempt to evolve the ~50 kbp *epoBCDEF* gene cluster from *Sorangium cellulosum* to produce epothilone in high yields in *E. coli.* Epothilones are an attractive class of anti-cancer compound that inhibit cell division by stabilizing the α/β-tubulin heterodimer.[40] Epothilone production has been achieved in *E. coli*,[40] paving the way for the generation of novel epothilone derivatives via genetic manipulation of the *epoBCDEF* cluster using the vast array of tools available in *E. coli*. However, the *epoBCDEF* enzyme complex expresses and folds poorly in *E. coli*, requiring extensive rational engineering and the expression of chaperones to produce trace amounts of epothilone. Without meaningful levels of epothilone production as a starting point, biosynthetic engineering in *E. coli* is not feasible.[41]

To evolve variants of the *epoBCDEF* cluster that are highly active in *E. coli*, we will couple epothilone production to recA1202 expression using a two-hybrid transcriptional system.[42] In this scheme, a split transcription factor would only be active if epothilone induces α/β-tubulin dimerization (**Figure 5.12**). We hypothesize that a directed evolution approach can evolve these enzymes to express, fold, and function properly in *E. coli* and produce enough epothilone to potentiate biosynthetic derivatization studies. If successful, we will unlock the ability to biosynthetically engineer the production of novel epothilone derivatives by harnessing the vast arsenal of genetic tools in *E. coli*.

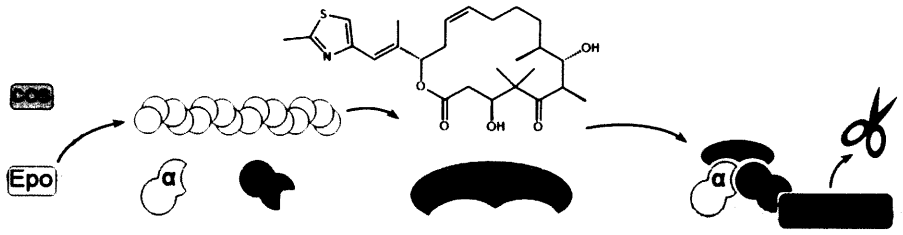

Figure 5.12 Selection circuit for evolving highly productive *epoBCDEF* in *E. coli*.
Successful epothilone production causes dimerization of the α-tubulin and β-tubulin domains of of a split transcription factor, which then induces recA1202 expression to package the cosmid encoding the successful *epoBCDEF* variant.

5.5 Conclusion

We have shown that it is feasible to use cosmids and helper lambda prophage as a vehicle for directed evolution. With further optimization, selection pressure should be easy to tune and monitor with moderate adjustments in incubation temperature and cheap optical monitoring. We expect that lambda-based continuous evolution will be highly generalizable and capable of evolving any BOI that can be coupled to recA1202 expression or cI cleavage. We envision adapting previously demonstrated M13 phage selection circuits to evolve proteases,[4] TALENs,[6] protein–protein interactions,[2] Cas9,[5] and more. In particular, this system is well-suited for evolving large BOIs that are difficult to access with current directed evolution methodology, such as assembly line enzymes that generate compounds of therapeutic interest.

5.6 Materials and Methods

General methods: All PCR reactions for generating recombineering targeting cassettes were performed using Q5 High Fidelity DNA Polymerase (New England BioLabs). Primers were obtained from Life Technologies and Sigma-Aldrich (**Table 5.4**). Colony PCR was performed with OneTaq® Quick-Load® 2X Master Mix (New England Biolabs). Site-directed mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent).

Lambda *red* recombineering: Recombineering was performed as previously described.[43,44] The following modifications (**Table 5.1**) were made to the CY2115 lysogen (see Appendix B.3 for prophage sequence) using the primers in **Table 5.4** and the R6K-kan-ccdB plasmid (GenBank Accession No. MH325106) to obtain the LseI lysogen:

Table 5.2 Modifications made to the CY2115 lysogen.

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
Mutating *rpsL* to confer streptomycin resistance	*rpsL(Str^R)*	N/A-Streptomycin resistance selected for directly using streptomycin plates.	rpsL.Forward and rpsL.Reverse (annealed oligos)	Making the genomic *rpsL* allele streptomycin resistant potentiates the use of *rpsL* counterselection for future engineering,[26] mutation rate assays,[27] or plasmid curing.
Deletion of *attR*	*ΔattR*	attR.kanccdB.Forward and attR.kanccdB.Reverse	delattR.Forward and delattR.Reverse (annealed oligos)	We temporarily deleted the attR site to prevent unwanted prophage excision[31] during lysogen modification.
Undoing the *Sam7* mutation	*S$^+$*	Sam7.kanccdB.Forward and Sam7.kanccdB.Reverse	holinS.Forward and holinS.Reverse (annealed oligos)	The *Sam7* mutation renders the prophage unable to lyse cells.[8] We repaired this mutation to allow efficiently release phage capsids without chloroform lysis.

Repairing the *stf* frameshift mutation	*stf*$^+$	stf.kanccdB.Forward and stf.kanccdB.Reverse	stf.Forward and stf.Reverse (annealed oligos)	Repair of the *stf* gene has been shown to increase the phage capsid adsorption rate by 8-fold.[30]
Deletion of *recA*	*ΔrecA*	recA.kanccdB.Forward and recA.kanccdB.Reverse	delrecA.Forward and delrecA.Reverse (annealed oligos)	We deleted the genomic allele of recA to prevent any unwanted cI cleavage.
Undoing the *attR* deletion	*attR*$^+$	attR.kanccdB.Forward and attR.kanccdB.Reverse	attR.Forward and attR.Reverse used to amplify from CY2115 genomic DNA	We undid the temporary attR deletion once we made all the desired modifications.

The following modifications (**Table 5.3**) were made to the BAC-cI857 plasmid (see Appendix B.3 for sequence) using the primers in **Table 5.4** and the R6K-kan-ccdB plasmid to obtain the different variants of the cI repressor:

Table 5.3 Modifications made to the BAC-cI857 plasmid.

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
Undoing the cI857 mutation	*cI*	cIwt.kanccdB.Forward and cIwt.kanccdB.Reverse	cIwt.Forward and cIwt.Reverse (annealed oligos)	Undoes the temperature-sensitive cI857 mutation to yield wild-type cI.
Mutating Gly112 to Glu112	*G112E*	G112E.kanccdB.Forward and G112E.kanccdB.Reverse	G112E.Forward and G112E.Reverse (annealed oligos)	Renders the cI repressor uncleavable.[14]

Table 5.4 Primers used for recombineering.

Primer Name	Sequence
rpsL.Forward	5′-acgaacacggcatactttacgcagcgcggagttcggcttccgcggcgtggtagtatatacacgagtacatacgccacgtt-3′
rpsL.Reverse	5′-aacgtggcgtatgtactcgtgtatatactaccacgccgcggaagccgaactccgcgctgcgtaaagtatgccgtgttcgt-3′
attR.kanccdB.Forward	5′-gcgacaggtttgatgacaaaaaattagcgcaagaagacaaaaatcaccttccctcatcagtgccaacatagtaag-3′
attR.kanccdB.Reverse	5′-cttaaaggtattaaaaacaacttttttgtctttttaccttcccgtttcgctccgctcattaggcgggc-3′
delattR.Forward	5′-gcgacaggtttgatgacaaaaaattagcgcaagaagacaaaaatcaccttagcgaaacgggaaggtaaaaagacaaaaagttgtttttaatacctttaag -3′
delattR.Reverse	5′-cttaaaggtattaaaaacaacttttttgtctttttaccttcccgtttcgctaaggtgatttttgtcttcttgcgctaatttttgtcatcaaacctgtcgc-3′

Sam7.kanccdB.Forward	5′-gtgcgtttacaaaaacagtaatcgacgcaacgatgtgcgccattatcgcccctgttgacaattaatcatcg gca-3′
Sam7.kancdB.Reverse	5′-taagcgagattgctacttagtccggcgaagtcgagaaggtcacgaatgaaccgctcattaggcgggc-3′
holinS.Forward	5′-ttacaaaaacagtaatcgacgcaacgatgtgcgccattatcgcctggttcattcgtgaccttctcgacttcg ccggactaagtagcaatct-3′
holinS.Reverse	5′-agattgctacttagtccggcgaagtcgagaaggtcacgaatgaaccaggcgataatggcgcacatcgtt gcgtcgattactgtttttgtaa-3′
stf.kanccdB.Forward	5′-aacgccaaaggcggttaaggtggtaatggatgaaacgaacagaaaagccccctgttgacaattaatca tcggca-3′
stf.kanccdB.Reverse	5′-tgagcgcggttggtgctgttggcgttccggtcagtgccggactgtccagtccgctcattaggcgggc-3′
stf.Forward	5′-caaaggcggttaaggtggtaatggatgaaacgaacagaaaagccccactggacagtccggcactga ccggaacgccaacagcaccaaccgc-3′
stf.Reverse	5′-gcggttggtgctgttggcgttccggtcagtgccggactgtccagtggggcttttctgttcgtttcatccattacca ccttaaccgcctttg-3′
recA.kanccdB.Forward	5′-atattgactatccggtattacccggcatgacaggagtaaaaatggctatcccctcatcagtgccaacatag taag-3′
recA.kanccdB.Reverse	5′-aaagggccgcagatgcgacccttgtgtatcaaacaagacgattaaaaatcccgctcattaggcgggc-3′
delrecA.Forward	5′-atattgactatccggtattacccggcatgacaggagtaaaaatggctatcgatttttaatcgtcttgtttgata cacaagggtcgcatctgcggcccttt-3′
delrecA.Reverse	5′-aaagggccgcagatgcgacccttgtgtatcaaacaagacgattaaaaatcgatagccatttttactcctgt catgccgggtaataccggatagtcaatat-3′
attR.Forward	5′-agacattcactacagttatggcgg-3′
attR.Reverse	5′-acgcgctggatatagaacgtattg-3′
clwt.kanccdB.Forward	5′-ctctggcgattgaagggctaaattcttcaacgctaactttgagaatttttcctgttgacaattaatcatcggc a-3′
clwt.kanccdB.Reverse	5′-gctttatttaatggcatcaatgcattaaatgcttataacgccgcattgctccgctcattaggcgggc-3′
clwt.Forward	5′-gggctaaattcttcaacgctaactttgagaatttttgcaagcaatgcggcgttataagcatttaatgcattgat gccatt-3′
clwt.Reverse	5′-aatggcatcaatgcattaaatgcttataacgccgcattgcttgcaaaaattctcaaagttagcgttgaagaat ttagccc-3′
G112E.kanccdB.Forward	5′-agtatgagtaccctgtttttttctcatgttcaggcacctgttgacaattaatcatcggca-3′
G112E.kanccdB.Reverse	5′-ttggtaaaggttctaagctcaggtgagaacatcccactccgctcattaggcgggc-3′
G112E.Forward	5′-agtatgagtaccctgtttttttctcatgttcaggcagaaatgttctcacctgagcttagaacctttaccaaagg-3′
G112E.Reverse	5′-cctttggtaaaggttctaagctcaggtgagaacatttctgcctgaacatgagaaaaaacagggtactcata ct-3′
cl.Forward	5′-ccagcagagaattaaggaaaacagacag-3′
cl.Reverse	5′-agttaaaaaaatctttcggcctgcatg-3′
psc101.kanccdB.Forward	5′-gccggcggcacctcgctaacggattcaccactccaagatgcataatgtgcccctcatcagtgccaacata gtaag-3′
psc101.kanccdB.Reverse	5′-tgacggagtagcatagggtttgcagaatccctgcttcgtccatttgacagccgctcattaggcgggc-3′
recA1202.Forward	5′-gcgggtaacctgaagaagtccaacacgctgc-3′
recA1202.Reverse	5′-gcagcgtgttggacttcttcaggttacccgc-3′

Reversing the *ind-1* mutation via plasmid cointegration: In order to reverse the *ind-1* mutation, we applied a plasmid cointegration strategy with the R6K-kan-ccdB-cI857 plasmid (see Appendix B.3 for sequence). A derivative of CY2115 (*rpsL(Str^R) ΔattR S$^+$ stf$^+$*) containing the psc101-gbaA recombineering plasmid was grown in LB containing 10 µg/mL tetracycline and 25 µg/mL chloramphenicol overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 µg/mL tetracycline and 25 µg/mL chloramphenicol and grown at 30 °C for ~2 h until attaining an OD_{600} of 0.3–0.4. The ccdA antitoxin was then induced by adding L-arabinose 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 R6K-kan-ccdB-cI857 plasmid (1.8 kV, 5.8 msec, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in super optimal broth with catabolite repression (SOC) with 2 mg/mL L-arabinose at 30 °C for 2 h, then plated on LB agar plates with 50 µg/mL kanamycin and 2 mg/mL L-arabinose and incubated for 24 h at 30 °C. Colonies that grew under these conditions had integrated the R6K-kan-ccdB-cI857 plasmid and several were picked and streaked on LB agar plates with 10 µg/mL tetracycline and 25 µg/mL chloramphenicol and without L-arabinose and incubated for 24 h at 30 °C. Without arabinose induction of the ccdA antitoxin, the ccdB toxin killed cells that have not excised the *kan^R-ccdB* cassette by resolving the cointegrant. The colonies that grew were screened by colony PCR to identify colonies that lost the *ind-1* mutation. Each colony was picked with a sterile 10 µL pipette tip and each swirled in a PCR tube containing 25 µL of 1× OneTaq® Quick-Load® 2X Master Mix (New England Biolabs) and 10 µM each of the primers cI.Forward and cI.Reverse (**Table 5.4**). The PCRs were performed using the following thermocycler PCR protocol and visualized by gel electrophoresis:

1. 94 °C 5 min
2. 94 °C 15 sec
3. 68 °C 15 sec
4. 54 °C 3 min
5. Return to step 2 for 30 cycles
6. 68 °C 5 min
7. 4 °C hold

The PCR reactions were then submitted for Sanger sequencing with the cI.Forward and cI.Reverse primers (**Table 5.4**).

Curing the psc101-gbaA recombineering plasmid: When we were unable to make recombineering modifications on the lysogen at certain points in the series of modifications, we suspected that the phage had duplicated by inserting into the recombineering plasmid. We were able to make modifications only after curing and then reintroducing the psc101-gbaA recombineering plasmid, presumably eliminating duplicate copies of the prophage. Normally the psc101-gbaA plasmid is cured by incubating the strain at 42 °C.[25] However, we were unable to cure the plasmid in this manner because the high temperature would induce the lambda prophage and kill the cells. Instead, we cured the psc101-gbaA plasmid by inserting a *kan^R+ccdB* cassette into the plasmid via recombineering, and then counterselecting against ccdB for cells that had lost the psc101-gbaA plasmid. In detail, a psc101-kanccdB targeting cassette was generated using the primers psc101.kanccdB.Forward and psc101.kanccdB.Reverse (**Table 5.4**) and the R6K-kan-ccdB template plasmid. The strain containing the psc101-gbaA recombineering plasmid to be cured was grown in LB containing 10 µg/mL tetracycline and 25 µg/mL chloramphenicol overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 µg/mL tetracycline and 25 µg/mL chloramphenicol 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 L-arabinose and L-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 psc101-kanccdB targeting cassette (1.8 kV, 5.8 msec, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in super optimal broth with catabolite repression (SOC) with 2 mg/mL L-arabinose at 30 °C for 2 h, then plated on LB agar plates with 50 µg/mL kanamycin and 2 mg/mL L-arabinose and incubated for 24 h at 30 °C. Colonies that grew under these conditions had integrated the *kan*^{R}+*ccdB* cassette into the psc101-gbaA plasmid and several were picked and streaked on LB agar plates with 25 µg/mL chloramphenicol and without L-arabinose or tetracycline, and incubated for 24 h at 30 °C. Without arabinose induction of the ccdA antitoxin, the ccdB toxin killed cells that have not lost the *kan*^{R}-*ccdB* cassette by curing the psc101-gbaA plasmid. The colonies that grew were screened for tetracycline sensitivity to confirm psc101-gbaA plasmid loss. Several colonies were picked with a sterile 10 µL pipette tip and each swirled in LB with 25 µg/mL chloramphenicol or LB with 10 µg/mL tetracycline and grown overnight at 30 °C (18–21 h). Colonies that only grew in LB with 25 µg/mL chloramphenicol and not in LB with 10 µg/mL tetracycline were cured.

Generation of pBAD-recA1202 via site-directed mutagenesis: Site-directed mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent), following the manufacturer's instructions. The recA1202 point mutation was introduced into the pBAD-recA plasmid (see Appendix B.3 for sequence) using the primers recA1202.Forward and recA1202.Reverse (**Table 5.4**).

LacZα activity assay for quantifying cI cleavage by recA: In order to determine the extent of derepression of the P_L promoter owing to cI cleavage by recA, the expression of the lacZα fragment from the P_L promoter at different temperatures and recA expression levels was measured via the cleavage of *o*-nitrophenyl-β-galactoside (*o*NPG) using an assay adapted from a previous publication.[45] In detail, the pL-lacZ reporter plasmid (see Appendix B.3 for sequence) was transformed into DH10B *E. coli* cells harboring different versions of the pBAD-recA plasmid and different versions of the BAC-cI plasmid. To obtain the data in **Figure 5.5** and **Figure 5.6**, each strain was grown in 200 µL LB with 10 µg/mL chloramphenicol, 100 µg/mL ampicillin, 50 µg/mL kanamycin, and 1 mM IPTG overnight at 32 °C shaking at 220 r.p.m. at 30 °C for 24 h in a parafilm-wrapped 96-well plate. Then 40 µL of each overnight was used to inoculate 1 mL of LB with 10 µg/mL chloramphenicol, 100 µg/mL ampicillin, 50 µg/mL kanamycin, and 1 mM IPTG, which was grown at 1000 r.p.m. at 30 °C for 3 h until the cultures reached mid-log phase. IPTG was added to induce the expression of the lacZω fragment from the DH10B genome that complements the lacZα fragment for the lacZ activity assay.

To obtain the recA1202 expression gradient data in **Figure 5.5**, 50 µL of each mid-log culture was distributed across 8 wells of a clear-bottomed 96-well plate, and 10 µL of a 10-fold dilution series of L-arabinose was added to each set of 8 wells to the following final concentrations: 0, 0.000111, 0.00111, 0.0111, 0.111, 1.11, 11.1 or 111 mM. The plate was wrapped in parafilm and incubated at 225 r.p.m. at 32 °C for 3 h.

To obtain the temperature gradient data in **Figure 5.6**, 60 µL of each mid-log culture was distributed across 8 wells of a PCR strip tube and incubated in a thermocycler with the following temperature gradient for 4 h: 30.0, 30.7, 32.2, 34.7, 37.6, 39.4, 41.0, and 42.0 °C.

After each incubation period at different L-arabinose concentrations or different temperatures, 60 µL of each culture was mixed with 90 µL of βgal mix (60 mM sodium

phosphate dibasic, 40 mM sodium phosphate monobasic, 10 mM potassium chloride, 1 mM magnesium sulfate, 26 mM 2-mercaptoethanol, 166 µg/mL egg-white lysozyme, 1.0 mg/mL oNPG, and 6.7% PopCulture lysis reagent) in a black, clear-bottomed 96-well plate. The OD_{600} and OD_{420} of each well was measured every 2 min over the course of 1 h in a Biotek Synergy H1 hybrid plate reader followed by double orbital shaking at 559 r.p.m. at 30 °C. The *o*NPG cleavage activity of each well was calculated by measuring the slope of the linear region of each OD_{420} trace (ΔOD_{420}/day), and dividing by the initial OD_{600} reading.

The data in **Figure 5.7** were collected in an experiment similar to the one described for **Figure 5.5** and **Figure 5.6**, except that a 6 mL mid-log culture of a strain carrying the pBAD-recA1202, BAC-cI857 and pL-lacZ plasmids was employed. 50 µL of a mid-log culture was distributed to 96 PCR tubes along with 10 µL of a 2-fold dilution series of L-arabinose to the final concentrations indicated in **Figure 5.7**. Then the PCR tubes were incubated in a thermocycler at the temperatures indicated in **Figure 5.7** for 4 h before analyzing lacZ activity, as described above.

The data in **Figure 5.8** were collected in experiments similar to the ones described for **Figure 5.5** and **Figure 5.6**, except that a 2-fold dilution series of arabinose was used instead of a 10-fold dilution series to yield the following final concentrations: 0, 0.430, 0.870, 1.70, 3.00, 6.90, 13.9, 27.8, 55.0, 111 mM.

Generating packaged cosmids: We generated packaged cosmids by adapting a previously reported protocol.[46] Briefly, a 2 mL LB culture with 10 µg/mL chloramphenicol and 100 µg/mL ampicillin was inoculated with a colony of LseI harboring the recA1202 cosmid (see Appendix B.3 for sequence) and grown overnight at 250 r.p.m. at 30 °C for 20 h. Next, 80 µL of the overnight was used to inoculate a fresh 2 mL LB culture with 10 µg/mL chloramphenicol and 100 µg/mL ampicillin, which grew to mid-log phase at 250 r.p.m. at 30 °C for 2.5 h. Next, 1mL of the mid-log culture was distributed evenly across 8 wells of a PCR strip-tube and incubated at 42 °C for 25 min in a thermocycler. Following that, the tubes were incubated at 37 °C for 5 h in the thermocycler. The 1 mL of media was then collected from the PCR tubes and recombined into a 1.5 microcentrifuge tube along with 20 µL of chloroform. The tube was shaken vigorously by hand for 10 min at room temperature, and then cellular debris was cleared by centrifugation at 10,000 r.p.m. for 1 min. The cleared supernatant containing the packaged cosmids was transferred to a 2 mL screw cap vial and stored at 4 °C.

Plaque assay: We performed plaque assays by adapting previously reported protocols.[46,47] A 2 mL LB culture with 10 µg/mL chloramphenicol and 100 µg/mL ampicillin was inoculated with the LseI strain and was grown overnight at 250 r.p.m. at 30 °C for 24 h. Next, 120 µL of overnight was used to inoculate a 3 mL LB culture with 10 µg/mL chloramphenicol, 100 µg/mL ampicillin, and 0.2% (w/v) maltose which was grown at 250 r.p.m. at 30 °C for 7 h to stationary phase. Maltose was added to the media to induce expression of the lamB receptor and ensure efficient phage particle adsorption.[48] 1 mL of the stationary phase culture was then pelleted at 10,000 r.p.m. for 1 min and resuspended in 500 µL of 10 mM magnesium sulfate. The stock of packaged recA1202 cosmid was diluted 1,000-fold in SM buffer with gelatin (100 mM sodium chloride, 8 mM magnesium sulfate, 50 mM Tris (pH 7.5), 0.002% (w/v) gelatin). Then 100 µL of diluted packaged recA1202 cosmid was incubated with 100 µL of LseI cells in magnesium sulfate at 30 °C

for 20 min. The cell/phage mixture was then added to 3 mL of molten top agarose (0.7% (w/v) agarose in LB with 0.2% (w/v) maltose) pre-equilibrated to 47 °C, mixed by inversion 5 times quickly, then poured onto a warm LB agar plate pre-equilibrated at 37 °C and distributed evenly by tilting the plate. The agarose overlay was allowed to harden at room temperature for 10 minutes, after which the plate was incubated at 40 °C overnight until a lawn with plaques appeared.

5.7 Acknowledgements
This work was supported by an NIH Director's New Innovator Award 1DP2GM119162 and the MIT School of Science Jarve Seed Fund for Innovation (both to M.D.S.). L.J.P. was supported by a National Science Foundation Graduate Research Fellowship under Grant No. 1122374.

5.8 References

[1] Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules, *Nature 472*, 499–503.

[2] Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., Wang, P., Malvar, T., and Liu, D. R. (2016) Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance, *Nature 533*, 58–63.

[3] Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. (2014) Negative selection and stringency modulation in phage-assisted continuous evolution, *Nat. Chem. Biol. 10*, 216–222.

[4] Dickinson, B. C., Packer, M. S., Badran, A. H., and Liu, D. R. (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations, *Nat. Commun. 5*, 5352.

[5] Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, *Nature 556*, 57–63.

[6] Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., Tsai, S. Q., Sander, J. D., Joung, J. K., and Liu, D. R. (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity, *Nat. Methods. 12*, 939–942.

[7] Hines, J. C., and Ray, D. S. (1980) Construction and characterization of new coliphage M13 cloning vectors, *Gene 11*, 207–218.

[8] Cronan, J. E. (2013) Improved plasmid-based system for fully regulated off-to-on gene expression in *Escherichia coli*: application to production of toxic proteins, *Plasmid 69*, 81–89.

[9] Little, J. W., and Michalowski, C. B. (2010) Stability and instability in the lysogenic state of phage lambda, *J. Bacteriol. 192*, 6064–6076.

[10] Tessman, E. S., and Peterson, P. (1985) Plaque color method for rapid isolation of novel recA mutants of *Escherichia coli* K-12: new classes of protease-constitutive recA mutants, *J. Bacteriol. 163*, 677–687.

[11] Badran, A. H., and Liu, D. R. (2015) Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra, *Nat. Commun. 6*, 8425.

[12] Chauthaiwale, V. M., Therwath, A., and Deshpande, V. V. (1992) Bacteriophage lambda as a cloning vector, *Microbiol. Rev. 56*, 577–591.

[13] Sussman, R., and Jacob, F. (1962) Sur un système de répression thermosensible chez le bactériophage L d'*Escherichia coli*, *Comptes Rendus Acad. Sci. 254*, 1517–1519.

[14] Gimble, F. S., and Sauer, R. T. (1986) λ repressor inactivation: properties of purified ind$^-$ proteins in the autodigestion and recA-mediated cleavage reactions, *J. Mol. Biol. 192*, 39–47.

[15] Little, J. W., Shepley, D. P., and Wert, D. W. (1999) Robustness of a gene regulatory circuit, *EMBO J. 18*, 4299–4307.

[16] Freifelder, D., and Levine, E. E. (1975) The formation of polylysogens during infection of *Escherichia coli* with bacteriophage λ, *Virology 63*, 428–437.

[17] Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proc. Natl. Acad. Sci. U.S.A. 97*, 6640–6645.
[18] Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol. Syst. Biol. 2*, 2006.0008.
[19] Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., and Wood, T. K. (2010) Cryptic prophages help bacteria cope with adverse environments, *Nat. Commun. 1*, 147.
[20] Sauer, R. T. (1978) DNA sequence of the bacteriophage lambda cI gene, *Nature 276*, 301–302.
[21] Donnenberg, M. S., and Kaper, J. B. (1991) Construction of an *eae* deletion nutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector, *Infect. Immun. 59*, 4310–4317.
[22] Gong, S., Yang, X. W., Li, C., and Heintz, N. (2002) Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication, *Genome Res. 12*, 1992–1998.
[23] Wang, J., Sarov, M., Rientjes, J., Fu, J., Hollak, H., Kranz, H., Xie, W., Stewart, A. F., and Zhang, Y. (2006) An improved recombineering approach by adding recA to λ red recombination, *Mol. Biotechnol. 32*, 43–54.
[24] Kolter, R., Inuzuka, M., and Helinski, D. R. (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K, *Cell 15*, 1199–1208.
[25] Wang, H., Bian, X., Xia, L., Ding, X., Muller, R., Zhang, Y., Fu, J., and Stewart, A. F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection, *Nucleic Acids Res. 42*, e37.
[26] Bird, A. W., Erler, A., Fu, J., Hériché, J. K., Maresca, M., Zhang, Y., Hyman, A. A., and Stewart, A. F. (2011) High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes, *Nat. Methods 9*, 103–109.
[27] Moore, C. L., Papa, L. J., III, and Shoulders, M. D. (2018) A processive protein chimera introduces mutations across defined DNA regions *in vivo*, *J. Am. Chem. Soc. 140*, 11560–11564.
[28] Hendrix, R. W., and Duda, R. L. (1992) Bacteriophage lambda PaPa: not the mother of all lambda phages, *Science 258*, 1145–1148.
[29] Haggård-Ljungquist, E., Halling, C., and Calendar, R. (1992) DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages, *J. Bacteriol. 174*, 1462–1477.
[30] Shao, Y., and Wang, I. N. (2008) Bacteriophage adsorption rate and optimal lysis time, *Genetics 180*, 471–482.
[31] Bushman, W., Thompson, J. F., Vargas, L., and Landy, A. (1985) Control of directionality in lambda site specific recombination, *Science 230*, 906–911.
[32] Feiss, M., Siegele, D. A., Rudolph, C. F., and Frackman, S. (1982) Cosmid DNA packaging in vivo, *Gene 17*, 123–130.

[33] Takeshi, M., and Kenichi, M. (1982) Identification of sequences necessary for packaging DNA into lambda phage heads, *Gene 20*, 267–279.
[34] Court, D. L., Oppenheim, A. B., and Adhya, S. L. (2007) A new look at bacteriophage lambda genetic networks, *J. Bacteriol. 189*, 298–304.
[35] Gimble, F. S., and Sauer, R. T. (1989) λ repressor mutants that are better substrates for recA-mediated cleavage, *J. Mol. Biol. 206*, 29–39.
[36] Winn, M., Fyans, J. K., Zhuo, Y., and Micklefield, J. (2016) Recent advances in engineering nonribosomal peptide assembly lines, *Nat. Prod. Rep. 33*, 317–347.
[37] Klaus, M., and Grininger, M. (2018) Engineering strategies for rational polyketide synthase design, *Nat. Prod. Rep. 35*, 1070–1081.
[38] Inoue, H., Nojima, H., and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids, *Gene 96*, 23–28.
[39] Siguret, V., Ribba, A. S., Chérel, G., Meyer, D., and Piétu, G. (1994) Effect of plasmid size on transformation efficiency by electroporation of *Escherichia coli* DH5α, *BioTechniques 16*, 422–426.
[40] Mutka, S. C., Carney, J. R., Liu, Y., and Kennedy, J. (2006) Heterologous production of epothilone C and D in *Escherichia coli*, *Biochemistry 45*, 1321–1330.
[41] Osswald, C., Zipf, G., Schmidt, G., Maier, J., Bernauer, H. S., Muller, R., and Wenzel, S. C. (2014) Modular construction of a functional artificial epothilone polyketide pathway, *ACS Synth. Biol. 3*, 759–772.
[42] Stynen, B., Tournu, H., Tavernier, J., and Van Dijck, P. (2012) Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system, *Microbiol. Mol. Biol. Rev. 76*, 331–382.
[43] Berman, C., Papa, L. J., III, Hendel, S. J., Moore, C. L., Suen, P. H., Weickhardt, A. F., Doan, N. D., Kumar, C. M., Uil, T. G., Butty, V. L., Hoeben, R. C., and Shoulders, M. D. (2018) An adaptable platform for directed evolution in human cells, *J. Am. Chem. Soc. 140*, 18093–18103.
[44] Papa, L. J., III, and Shoulders, M. D. (2019) Genetic Engineering by DNA Recombineering, *Curr. Protoc. Chem. Biol. 11*, e70.
[45] Schaefer, J., Jovanovic, G., Kotta-Loizou, I., and Buck, M. (2016) Single-step method for beta-galactosidase assays in Escherichia coli using a 96-well microplate reader, *Anal. Biochem. 503*, 56–57.
[46] Cronan, J. E. (2003) Cosmid-based system for transient expression and absolute off-to-on transcriptional control of *Escherichia coli* genes, *J. Bacteriol. 185*, 6522–6529.
[47] Sambrook, J., and Russell, D. W. (2006) Plating bacteriophage λ, *Cold Spring Harb. Protoc. 2006*, pdb.prot3961.
[48] Schwartz, M. (1976) The adsorption of coliphage lambda to its host: effect of variations in the surface density of receptor and in phage-receptor affinity, *J. Mol. Biol. 103*, 521–536.

Chapter 6: Perspectives and Future Directions

6.1 Perspective of the field

Fully or partially *in vitro* directed evolution campaigns only vaguely resemble the miraculous process that has created those who conduct such experiments. One could argue that the misguided imposition of artificial reason and order—which are inherent to human nature and experimental design—upon directed evolution workflows stifles a process that has succeeded over billions of years only through chaos and lack of restraint. When researchers try to "evolve" a new ability, they almost always start with or focus on one gene rather than large sets of genes as nature does. In an effort to expedite the search for an improved variant, we often focus mutations on particular active sites or residues we think are important and neglect the important regions we could not possibly know about or understand. In our impatience, we use very high selection pressure to obtain variants in one discrete step, favoring mutants that make rare evolutionary leaps with one or two mutations at the cost of more gradual multi-mutation evolutionary trajectories that might lead to far greater activities. In the fitness landscape of all possible variants, how often do we trap ourselves on a measly one-to-two mutation hilltop within an evolutionary mountain range? There could be benefits to lower selection pressure applied over many generations compared to sudden bursts of selection followed by mutagenic frenzies in a discontinuous directed evolution experiment.

Natural evolution has had the benefit of countless generations of countless species in a wide range of changing environments to produce the marvelous solutions we admire. We have a lot less time and space in the laboratory, so it is a challenge to create an artificial environment that can truly replicate natural evolution. Continuous *in vivo* evolution is the most practical way to explore massive libraries and long trajectories within a reasonable time frame. There are several notable examples in which continuous *in vivo* evolution methods worked beautifully, with the evolution experiment itself appearing to require minimal labor.[1-16] However, a huge amount of effort and pre-optimization was required in constructing the genetic circuits for each experiment, which restricts continuous evolution to the expertise of the labs that focus on the methodology. One need not be an expert in bioreactors, *in vivo* mutagenesis, and artificial genetic circuits to apply discontinuous mutagenesis to one's favorite biomolecule. Until the methodology is developed and generalized further, the benefits of continuous evolution—including larger libraries, more generations, the potential for large scales and parallelization, and access to more trajectories[17,18]—will be lost to the vast majority of laboratories and engineering efforts.

6.2 Future directions

6.2.1 Improving adenovirus-based directed evolution in human cells

In Chapter 2, I outline the difficulties we encountered engineering the large adenoviral genome using lambda *red* recombineering. While challenging, recombineering was the most practical and precise way to generate and test many variants of the adenoviral genome with different genes trans-complemented, such as AdProt, AdPol, pTP, fiber, PVI, and DBP. Fortunately, once we finished developing the AdProt-based system, we were able to make a finalized, user-friendly version of the adenoviral genome in which any target biomolecules of interest can be inserted into a gateway cassette without the need for recombineering expertise. In Chapter 3, we demonstrate the evolution of tTA by coupling its activity to AdProt expression and we show that an AdProt inhibitor can be used to tune selection pressure. Features such as the gateway-compatible adenoviral vector and the AdProt inhibitor to tune selection pressure are great examples of how to make the system more generalizable and easier to apply. Future efforts should focus on streamlining the methodology and making it more robust for adoption by other laboratories.

One potential improvement we can make is to re-engineer the adenoviral genome we are using to shorten the viral life cycle and accelerate evolution. The adenoviruses used in our current platform are missing several genes, some of which are trans-complemented and some of which are not. The removal of these genes from the adenoviral genome appears to lengthen the life cycle and/or decrease the yield of viruses relative to the wild-type virus. As we discovered in our attempts to evolve β-catenin in Appendix A, ΔAdpolΔAdProt viruses that reacquired the E1 genes from the host cell replicated much faster than their E1-deleted progenitors. While it is unclear if the effect is due to the reacquisition of the E1 genes or the loss of the β-catenin cassette, restoring the E1 genes could potentially allow us to perform more passages in less time. Additionally, the E3 deletion in our current viruses eliminates several non-essential genes,[19] including the adenoviral death protein (ADP) that is responsible for efficient cell lysis and viral release.[20] Reinserting ADP into an E3-deleted adenovirus has been shown to accelerate adenovirus propagation,[21] and could shorten the time required for each passage. While reinserting the E1 and ADP genes would decrease the BOI cargo capacity of the genome, we could compensate by expanding the E3 deletion to remove more unnecessary portions of the genome[22] and use a smaller fluorescent protein such as iLov,[23] thus leaving enough room for 4–5 kb BOIs.

Another important improvement we could make is to accelerate the viral rescue process, which is one of the main bottlenecks in our workflow. Currently, it takes about ~20 days to obtain an initial population of infectious BOI-carrying virus after transfecting the *E. coli*-derived adenoviral plasmid. Adenoviral rescue from transfected DNA vectors is renowned for being an inefficient process.[24] One widely suspected reason for the inefficiency is that *E. coli* derived DNA lacks the covalently linked terminal proteins at the ends of the linear genome, which serve as the primers of replication during adenovirus replication.[24,25] Rescue efficiency can be improved by overexpressing the terminal protein in the transfected cell lines,[26,27] or recombining genes of interest into purified adenoviral genomes that contain the native covalently linked terminal proteins.[24,25] These strategies could help us shorten the wait time for viral rescue, but it would be ideal to avoid the need

for viral rescue from naked DNA altogether. An attractive alternative for producing the necessary viruses is *in vivo* homologous recombination with live replicating adenovirus in human cells.

Others have shown that recombinant virus can be efficiently produced in just a few days by first infecting cells with a ΔAdProt parental adenovirus, and then transfecting in a cassette with AdProt and a gene encoding a BOI that will recombine with the replicating virus. Recombinant BOI- and AdProt-containing progeny virus can then be selected for on cells that do not trans-complement AdProt, thus using AdProt as a positive selection marker for recombination.[28] Since we use AdProt trans-complementation as our directed evolution selection couple, we would need to use a different positive selection marker for viral recombination. As mentioned above, if we switch to E1-containing adenoviruses in which the BOI and a fluorescent marker are inserted to the right of the E4 region, we could potentially use the E4 region as a positive selection marker for recombination, since the E4 genes have been successfully trans-complemented before.[29] Thus, we could construct our next generation adenoviruses by recombining an E4-BOI-fluorescent marker cassette with ΔE4 virus in an E4 trans-complementing cell line, then selecting for recombinant fluorescent plaques on E4-lacking cells (**Figure 6.1**).

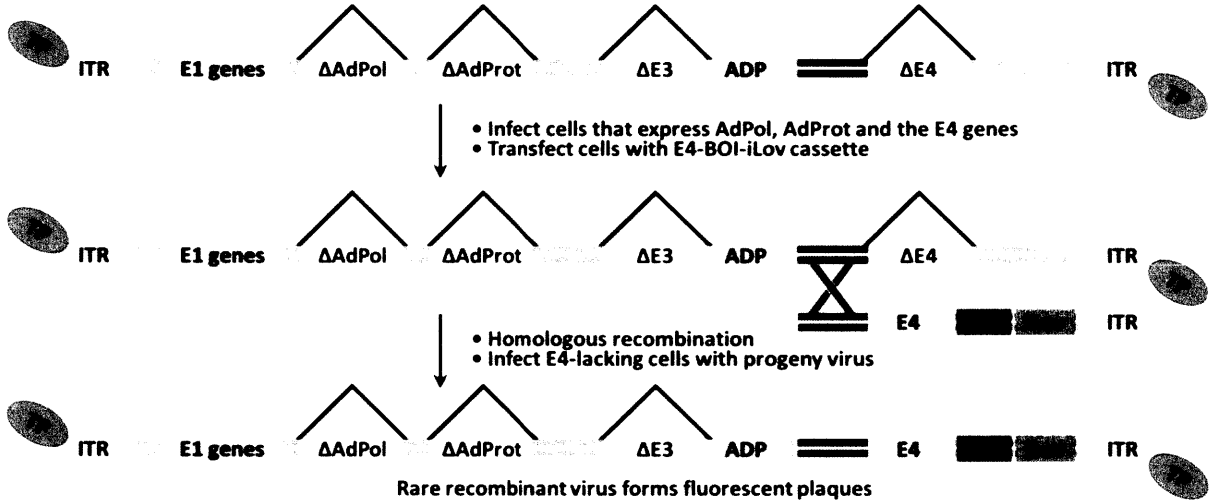

Figure 6.1 Production of recombinant virus using E4 selection.
By recombining an E4-BOI-iLov cassette with live replicating ΔE4 virus, and selecting for recombination on E4-lacking cells, we can produce BOI-containing adenoviruses in a few days rather than a few weeks.

Along with streamlining the construction of recombinant adenovirus, we could improve the system by making the construction of selection cell lines simpler. Currently we produce selection cell lines by integrating the selection circuit into the host genome by stable plasmid transfection or lentiviral transduction. These methods produce a heterogeneous population of cells in which the circuit has integrated into different locations of the genome in different cells, resulting in variable circuit expression levels. Sometimes it is desirable to isolate single clones of this population to get reproducible

selection circuit expression, which adds to the time and effort required for performing an evolution experiment. We can decrease the time needed to produce clonal selection cell lines by employing a previously reported "landing pad" strategy.[30] In this strategy, the selection circuit would be integrated into a specific genomic locus in a pre-inserted "landing pad" using a site specific recombinase.

Other general improvements to the platform could include employing suspension cell cultures, developing a negative selection strategy, and exploring other selection genes. Switching from adherent cell culture to suspension cell culture would dramatically increase the scale of the evolution experiments we could perform and allow us to explore more variants. Suspension cell culture also simplifies cell line passaging by eliminating the need to trypsinize cell monolayers. Negative selection is required for evolving highly specific new activities by eliminating residual parent activity. We could accomplish negative selection by coupling BOI activity to the expression of a gene that is toxic to the host, such as thymidine kinase.[31] Finally, while AdProt was the first selection gene we have developed as a selection transgene, there are various other adenoviral genes that have been successfully trans-complemented and could have different advantages as selection transgenes, such as fiber,[32-34] pTP,[35] DBP,[36,37] AdPol,[38] and 100K.[39]

6.2.2 Expanding the MutaT7 mutational spectrum and target species

In Chapter 4, we describe how to make highly focused C→T and A→G mutations with MutaT7. In order to achieve high mutagenicity with minimal toxicity, we found that we needed to tightly control expression of MutaT7, with only very low levels of MutaT7 expression inducing high rates of mutagenesis. Currently, the rApo1–T7 fusion induces an optimal rate of mutagenesis in the fully repressed state from the weak $P_{A1LacO\text{-}Tenth}$ promoter. In order to render the rApo1–T7 strain truly inducible, the $P_{A1LacO\text{-}Tenth}$ promoter likely needs to be engineered further to completely eliminate basal expression. On the other hand, the tadA–T7 fusion had lower activity, with no significant mutagenesis observed in the fully repressed state, and lower activity than rApo1–T7 fusion upon IPTG induction. We suspect that an ideal iteration of this system entails tightly controlled, minimal MutaT7 expression, and very high DNA damaging enzyme activity. Therefore, future efforts to optimize the system in *E. coli* should focus on optimizing the strength and repressibility of the $P_{A1LacO\text{-}Tenth}$ promoter, and on improving the activity of the DNA damaging enzymes, either through linker optimization or by improving catalytic efficiency. Fortunately, more active variants of the rApo1 and CDA1 cytidine deaminases have been recently developed through continuous directed evolution and could readily be applied to the current system.[12] The tadA adenosine deaminase could also likely be improved through directed evolution approaches. Furthermore, in order to expand the mutation spectrum beyond C→T and A→G base changes, new DNA damaging enzymes will likely need to be evolved or discovered. The development of tadA deoxyadenosine deaminase[40] serves as an excellent example for how to create such enzymes from of the vast repertoire of RNA modifying enzymes[41] that can potentially be reprogrammed to attack DNA.

Another major direction for future work with MutaT7 will be applying it in a wide variety of organisms. The broad activity of the ugi protein inhibitor against many homologs of ung[42] and the similarly broad activity of T7 RNA polymerase across diverse organisms[43-48] suggests that the system should be highly portable. MutaT7 will be particularly useful for boosting the mutations rates and library sizes in our adenovirus-based evolution system in human cells. The ugi inhibitor[42] and T7 RNA polymerase both function in human cells,[43] but will likely require nuclear localization signals[49] for optimal activity against adenoviral genomes replicating in the nucleus. Additionally, MutaT7 expression levels will likely need to be carefully engineered as they were in E. coli. MutaT7 may need to be integrated into the human genome and tightly repressed for optimal control of activity. In order to focus mutations on the adenovirus-borne GOI, a T7 promoter will need to be placed upstream. Preventing downstream mutations may be complicated, because it is unclear whether a highly repetitive downstream terminator array will be genetically stable within the adenoviral genome. Furthermore, the terminator array could potentially interfere with adenoviral replication owing to hairpin formation. An alternative approach to limiting unwanted downstream mutations might be to place the T7 promoter-GOI cassette at one end of the genome with the T7 promoter pointing away from the rest of the adenoviral genome. Because the adenoviral genome is linear, the T7 RNA polymerase will simply run off the end of the genome. We expect that this arrangement will permit mutations rates far above the error catastrophe limit for adenovirus.

6.2.3 Optimizing and applying lambda-based directed evolution

While the lambda-based directed evolution system described in Chapter 5 is still in early development, we expect that this system will greatly facilitate continuous directed evolution efforts in *E. coli*. Future work will focus on applying the system to the evolution of T7 RNA polymerases with altered sequence specificity and to epothilone synthetases with high activity in *E. coli*. Another exciting goal will be to optimize the system for use in turbidostat bioreactors. Because lambda prophage induction causes cell lysis and a reduction in culture turbidity, a turbidostat bioreactor can be adapted to maintain constant selection pressure by monitoring culture turbidity and modulating the dilution rate of cells fed into the reactor and/or increasing the reactor temperature. The automation that bioreactors provide should make it easy to perform many experiments in parallel. Large fermenters should also enable experiments on the order of several liters, thus easily exploring trillions of variants assuming one BOI-cosmid variant per cell.

Another major benefit of the system is the depth of the literature describing lambda phage and its countless variants. Many aspects of the phage and its life cycle—from lysis time,[50] to receptor affinity,[51,52] to lysogen stability, [53-55] to temperature sensitivity[56]—can be tuned with well-characterized mutations. As one of the first and most widely used genetic models, lambda phage is indisputably one of the best understood and most engineerable organisms. While we cannot anticipate every issue that might arise in the development of this system, there is likely a lambda phage mutant that can resolve it.

6.3 References

[1] Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules, *Nature 472*, 499–503.

[2] Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., and Liu, D. R. (2013) Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution, *Proc. Natl. Acad. Sci. U.S.A. 110*, 9007–9012.

[3] Dickinson, B. C., Packer, M. S., Badran, A. H., and Liu, D. R. (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations, *Nat. Commun. 5*, 5352.

[4] Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. (2014) Negative selection and stringency modulation in phage-assisted continuous evolution, *Nat. Chem. Biol. 10*, 216–222.

[5] Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., Tsai, S. Q., Sander, J. D., Joung, J. K., and Liu, D. R. (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity, *Nat. Methods 12*, 939–942.

[6] Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., Wang, P., Malvar, T., and Liu, D. R. (2016) Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance, *Nature 533*, 58–63.

[7] Packer, M. S., Rees, H. A., and Liu, D. R. (2017) Phage-assisted continuous evolution of proteases with altered substrate specificity, *Nat. Commun. 8*, 956.

[8] Bryson, D. I., Fan, C., Guo, L. T., Miller, C., Soll, D., and Liu, D. R. (2017) Continuous directed evolution of aminoacyl-tRNA synthetases, *Nat. Chem. Biol. 13*, 1253–1260.

[9] Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, *Nature 556*, 57–63.

[10] Wang, T., Badran, A. H., Huang, T. P., and Liu, D. R. (2018) Continuous directed evolution of proteins with improved soluble expression, *Nat. Chem. Biol. 14*, 972–980.

[11] Roth, T. B., Woolston, B. M., Stephanopoulos, G., and Liu, D. R. (2019) Phage-assisted evolution of *Bacillus methanolicus* methanol dehydrogenase 2, *ACS Synth. Biol. 8*, 796–806.

[12] Thuronyi, B. W., Koblan, L. W., Levy, J. M., Yeh, W. H., Zheng, C., Newby, G. A., Wilson, C., Bhaumik, M., Shubina-Oleinik, O., Holt, J. R., and Liu, D. R. (2019) Continuous evolution of base editors with expanded target compatibility and improved activity, *Nat. Biotechnol. 37*, 1070–1079.

[13] Marzio, G., Verhoef, K., Vink, M., and Berkhout, B. (2001) In vitro evolution of a highly replicating, doxycycline-dependent HIV for applications in vaccine studies, *Proc. Natl. Acad. Sci. U.S.A. 98*, 6342–6347.

[14] Zhou, X., Vink, M., Klaver, B., Berkhout, B., and Das, A. T. (2006) Optimization of the Tet-On system for regulated gene expression through viral evolution, *Gene Ther. 13*, 1382–1390.

[15] Verhoef, K., Marzio, G., Hillen, W., Bujard, H., and Berkhout, B. (2001) Strict control of human immunodeficiency virus type 1 replication by a genetic switch: Tet for Tat, *J. Virol. 75*, 979–987.
[16] Das, A. T., Zhou, X., Vink, M., Klaver, B., Verhoef, K., Marzio, G., and Berkhout, B. (2004) Viral evolution as a tool to improve the tetracycline-regulated gene expression system, *J. Biol. Chem. 279*, 18776–18782.
[17] Badran, A. H., and Liu, D. R. (2015) *In vivo* continuous directed evolution, *Curr. Opin. Chem. Biol. 24*, 1–10.
[18] Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins, *Nat. Rev. Genet. 16*, 379–394.
[19] Graham, F. L., and Prevec, L. (1991) Manipulation of adenovirus vectors, *Methods Mol. Biol. 7*, 109–128.
[20] Doronin, K., Toth, K., Kuppuswamy, M., Krajcsi, P., Tollefson, A. E., and Wold, W. S. (2003) Overexpression of the ADP (E3-11.6K) protein increases cell lysis and spread of adenovirus, *Virology 305*, 378–387.
[21] Tollefson, A. E., Ying, B., Doronin, K., Sidor, P. D., and Wold, W. S. (2007) Identification of a new human adenovirus protein encoded by a novel late I-strand transcription unit, *J. Virol. 81*, 12918–12926.
[22] Ranheim, T. S., Shisler, J., Horton, T. M., Wold, L. J., Gooding, L. R., and Wold, W. S. (1993) Characterization of mutants within the gene for the adenovirus E3 14.7-kilodalton protein which prevents cytolysis by tumor necrosis factor, *J. Virol. 67*, 2159–2167.
[23] Kemp, V., Dautzenberg, I. J. C., Cramer, S. J., Hoeben, R. C., and van den Wollenberg, D. J. M. (2018) Characterization of a replicating expanded tropism oncolytic reovirus carrying the adenovirus E4orf4 gene, *Gene Ther. 25*, 331–344.
[24] Choi, E. W., Seen, D. S., Song, Y. B., Son, H. S., Jung, N. C., Huh, W. K., Hahn, J. S., Kim, K., Jeong, J. Y., and Lee, T. G. (2012) AdHTS: a high-throughput system for generating recombinant adenoviruses, *J. Biotechnol. 162*, 246–252.
[25] Miyake, S., Makimura, M., Kanegae, Y., Harada, S., Sato, Y., Takamori, K., Tokuda, C., and Saito, I. (1996) Efficient generation of recombinant adenoviruses using adenovirus DNA-terminal protein complex and a cosmid bearing the full-length virus genome, *Proc. Natl. Acad. Sci. U.S.A. 93*, 1320–1324.
[26] Wu, N., Zhang, H., Deng, F., Li, R., Zhang, W., Chen, X., Wen, S., Wang, N., Zhang, J., Yin, L., Liao, Z., Zhang, Z., Zhang, Q., Yan, Z., Liu, W., Wu, D., Ye, J., Deng, Y., Yang, K., Luu, H. H., Haydon, R. C., and He, T. C. (2014) Overexpression of Ad5 precursor terminal protein accelerates recombinant adenovirus packaging and amplification in HEK-293 packaging cells, *Gene Ther. 21*, 629–637.
[27] Lee, D., Liu, J., Junn, H. J., Lee, E. J., Jeong, K. S., and Seol, D. W. (2019) No more helper adenovirus: production of gutless adenovirus (GLAd) free of adenovirus and replication-competent adenovirus (RCA) contaminants, *Exp. Mol. Med. 51*, 127.
[28] Elahi, S. M., Oualikene, W., Naghdi, L., O'Connor-McCourt, M., and Massie, B. (2002) Adenovirus-based libraries: efficient generation of recombinant adenoviruses by positive selection with the adenovirus protease, *Gene Ther. 9*, 1238-1246.
[29] Yeh, P., Dedieu, J. F., Orsini, C., Vigne, E., Denefle, P., and Perricaudet, M. (1996) Efficient dual transcomplementation of adenovirus E1 and E4 regions from a 293-derived cell line expressing a minimal E4 functional unit, *J. Virol. 70*, 559–565.

[30] Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W. A., Lowe, A., Das, J., Bandara, K., Baijuraj, S., Summers, N. M., Lu, T. K., Zhang, L., and Weiss, R. (2018) A multi-landing pad DNA integration platform for mammalian cell engineering, *Nucleic Acids Res. 46*, 4072–4086.

[31] Black, M. E., Newcomb, T. G., Wilson, H. M., and Loeb, L. A. (1996) Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy, *Proc. Natl. Acad. Sci. U.S.A. 93*, 3525–3529.

[32] Legrand, V., Spehner, D., Schlesinger, Y., Settelen, N., Pavirani, A., and Mehtali, M. (1999) Fiberless recombinant adenoviruses: virus maturation and infectivity in the absence of fiber, *J. Virol. 73*, 907–919.

[33] Von Seggern, D. J., Chiu, C. Y., Fleck, S. K., Stewart, P. L., and Nemerow, G. R. (1999) A helper-independent adenovirus vector with E1, E3, and fiber deleted: structure and infectivity of fiberless particles, *J. Virol. 73*, 1601–1608.

[34] Von Seggern, D. J., Kehler, J., Endo, R. I., and Nemerow, G. R. (1998) Complementation of a fibre mutant adenovirus by packaging cell lines stably expressing the adenovirus type 5 fibre protein, *J. Gen. Virol. 79* 1461–1468.

[35] Hodges, B. L., Serra, D., Hu, H., Begy, C. A., Chamberlain, J. S., and Amalfitano, A. (2000) Multiply deleted [E1, polymerase-, and pTP-] adenovirus vector persists despite deletion of the preterminal protein, *J. Gene Med. 2*, 250–259.

[36] Zhou, H., O'Neal, W., Morral, N., and Beaudet, A. L. (1996) Development of a complementing cell line and a system for construction of adenovirus vectors with E1 and E2a deleted, *J. Virol. 70*, 7030–7038.

[37] Rice, S. A., and Klessig, D. F. (1985) Isolation and analysis of adenovirus type 5 mutants containing deletions in the gene encoding the DNA-binding protein, *J. Virol. 56*, 767–778.

[38] Uil, T. G., Vellinga, J., de Vrij, J., van den Hengel, S. K., Rabelink, M. J., Cramer, S. J., Eekels, J. J., Ariyurek, Y., van Galen, M., and Hoeben, R. C. (2011) Directed adenovirus evolution using engineered mutator viral polymerases, *Nucleic Acids Res. 39*, e30.

[39] Hodges, B. L., Evans, H. K., Everett, R. S., Ding, E. Y., Serra, D., and Amalfitano, A. (2001) Adenovirus vectors with the 100K gene deleted and their potential for multiple gene therapy applications, *J. Virol. 75*, 5913–5920.

[40] Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage, *Nature 551*, 464–471.

[41] Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017) Dynamic RNA modifications in gene expression regulation, *Cell 169*, 1187–1200.

[42] Wang, Z., and Mosbaugh, D. W. (1989) Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase, *J. Biol. Chem. 264*, 1163–1171.

[43] Meyer-Ficca, M. L., Meyer, R. G., Kaiser, H., Brack, A. R., Kandolf, R., and Kupper, J. H. (2004) Comparative analysis of inducible expression systems in transient transfection studies, *Anal. Biochem. 334*, 9–19.

[44] Conrad, B., Savchenko, R. S., Breves, R., and Hofemeister, J. (1996) *A T7 promoter-specific, inducible protein expression system for* Bacillus subtilis, *Mol. Gen. Genet. 250*, 230–236.

[45] McBride, K. E., Schaaf, D. J., Daley, M., and Stalker, D. M. (1994) 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.
[46] Lieber, A., Sandig, V., and Strauss, M. (1993) A mutant T7 phage promoter is specifically transcribed by T7-RNA polymerase in mammalian cells, *Eur. J. Biochem. 217*, 387–394.
[47] Weinstock, M. T., Hesek, E. D., Wilson, C. M., and Gibson, D. G. (2016) *Vibrio natriegens* as a fast-growing host for molecular biology, *Nat. Methods 13*, 849–851.
[48] Dower, K., and Rosbash, M. (2002) T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export, *RNA 8*, 686–697.
[49] Lieber, A., Kiessling, U., and Strauss, M. (1989) High level gene expression in mammalian cells by a nuclear T7-phase RNA polymerase, *Nucleic Acids Res. 17*, 8485–8493.
[50] Dennehy, J. J., and Wang, I. N. (2011) Factors influencing lysis time stochasticity in bacteriophage lambda, *BMC Microbiol. 11*, 174.
[51] Werts, C., Michel, V., Hofnung, M., and Charbit, A. (1994) Adsorption of bacteriophage lambda on the LamB protein of *Escherichia coli* K-12: point mutations in gene J of lambda responsible for extended host range, *J. Bacteriol. 176*, 941–947.
[52] Gallet, R., Kannoly, S., and Wang, I. N. (2011) Effects of bacteriophage traits on plaque formation, *BMC Microbiol. 11*, 181.
[53] Little, J. W., and Michalowski, C. B. (2010) Stability and instability in the lysogenic state of phage lambda, *J. Bacteriol. 192*, 6064–6076.
[54] Little, J. W., Shepley, D. P., and Wert, D. W. (1999) Robustness of a gene regulatory circuit, *EMBO J. 18*, 4299–4307.
[55] Gimble, F. S., and Sauer, R. T. (1989) λ repressor mutants that are better substrates for recA-mediated cleavage, *J. Mol. Biol. 206*, 29–39.
[56] Lieb, M. (1979) Heat-sensitive lambda repressors retain partial activity during bacteriophage induction, *J. Virol. 32*, 162–166.

Appendix A: β-catenin Directed Evolution

A.1 Author Contributions

Louis J. Papa constructed the pLVX.TCF.AdProt vector and the adenoviral vectors. L.J.P. performed the directed evolution of β-catenin and analyzed the resulting viruses. Samuel J. Hendel stably integrated the pLVX.TCF.AdProt vector into GH329 cells. Matthew D. Shoulders acquired funding and supervised experiments and data analysis.

A.2 Introduction

After successfully evolving doxycycline-resistant variants of the tTA transcription factor in our adenovirus-based system, one objective we aim to pursue is study of the evolution of a transcription factor that is more relevant to human health. We also sought to evolve a biomolecule that functions within a network that is absent from *E. coli* and yeast, and thus cannot be easily evolved in those contexts. We turned to β-catenin, a transcription co-factor that is activated by Wnt signaling, and forms a complex with TCF4 or other similar transcription co-factors to express genes involved with proliferation, such as c-myc.[1] β-catenin is constitutively degraded in the absence of Wnt signaling and is maintained at low levels and does not activate the transcription of its target genes. Upon upstream Wnt signaling, β-catenin degradation decreases, and β-catenin begins to accumulate and activate the transcription of its target genes.[2] However, oncogenic mutations, particularly in the N-terminal tail, allow β-catenin to escape this regulation, leading to β-catenin accumulation in the absence of Wnt signaling, aberrant transcription, and cancer.[2-4] β-catenin is a potential drug target for the treatment of certain cancers with several promising inhibitors reported in the literature,[5-9] although no direct inhibitors of β-catenin have yet been approved to treat cancer.[10]

Despite the importance of drug-resistance in the treatment of cancer, the ability of β-catenin to evolve resistance to inhibitors is largely unknown. Our platform provides a means to rapidly assess whether drug-resistant mutational escape pathways exist, much as is already done in the RNA virus field for antiviral drug leads,[11-13] and inform efforts to generate more effective therapeutics targeting β-catenin. We would also like to use the platform to explore how chaperones affect the drug-resistance pathways accessible to β-catenin. Our group and others have shown that the evolutionary pathways available to proteins can depend heavily on the capacity of the proteostasis network to buffer the negative effects that mutations can have on protein folding,[14-20] suggesting that the modulation of chaperones could be a potential therapeutic strategy for the prevention of drug resistance. We hypothesize that chaperones, such as HSP90, could play an important role in potentiating the evolution of β-catenin drug resistance, especially for inhibitors such as carnosic acid, which is thought to promote β-catenin degradation by inducing aggregation.[21] For example, β-catenin could conceivably acquire mutations that allow it to more effectively utilize chaperones and lessen the aggregating effect of carnosic acid, but decreasing chaperone capacity could cause such drug-resistance mutations to become ineffective. β-catenin could also conceivably acquire mutations that cause it to no longer bind an inhibitor, but that are slightly destabilizing. Decreasing chaperone capacity could render such mutations highly destabilizing and thus decrease the number of accessible drug-resistance mutations.

By coupling AdProt expression to β-catenin accumulation, we aimed to select for oncogenic variants of β-catenin that evade normal regulation through degradation using our platform. Presumably, variants of β-catenin that cause sufficient accumulation will induce more AdProt expression and thus produce more infectious virions, conferring a selective advantage to the adenovirus carrying that variant. We can further select for drug-resistant oncogenic variants by applying increasing amounts of drug during evolution once oncogenic variants have appeared.

Initial attempts to evolve a TetR–β-catenin fusion protein, which would presumably induce transcription from the tTA operator already in our existing selector cells, failed to

yield oncogenic mutations that fixed in the population. Interestingly, we found that in many of the attempted evolutions, the adenovirus seemed to have undergone a recombination event that both replaced the TetR–β-catenin fusion protein with the endogenous adenoviral E1 genes and seemed to confer a selective advantage to the viruses that had undergone the recombination. It was unclear if the TetR–β-catenin fusion was even functional and whether the E1 recombination event prevented the success of the evolution or not. To obviate both issues in future evolutions, we constructed a selection circuit that should be inducible by natural, unfused β-catenin, and inserted it into a cell line that prevents the E1 recombination event. Efforts to characterize and optimize this circuit and to evolve β-catenin are ongoing.

A.3 Results

In an initial attempt to couple β-catenin accumulation to AdProt, we fused β-catenin to TetR to create the tBcat fusion protein, which can theoretically bind and activate the tTA operator that is already present in our selector cells (**Figure A.1**). β-catenin was shown previously to tolerate N-terminal fusion of Lef1, one of its transcriptional co-factor partners, and activate transcription as a fusion protein.[22]

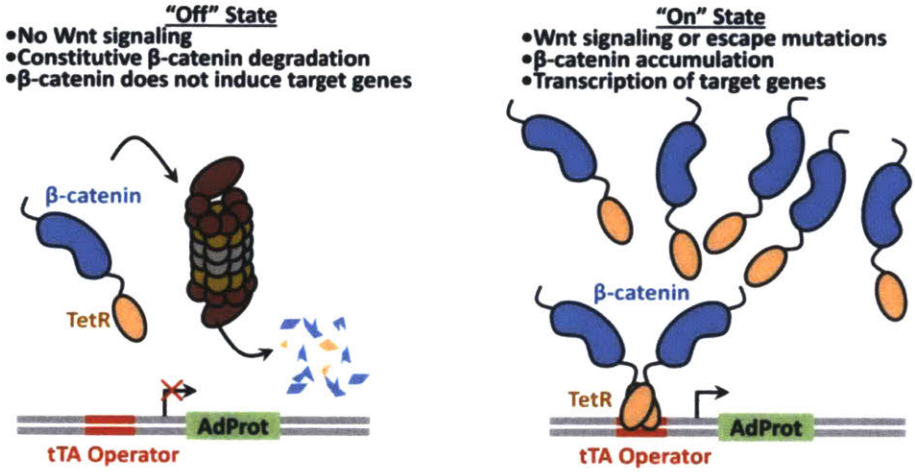

Figure A.1 Selection circuit utilizing a TetR–β-catenin fusion. When β-catenin signaling is in the off state, tBcat is constitutively degraded and AdProt is thus not expressed (*left*). When either the Wnt/β-catenin signaling pathway is activated, or β-catenin acquires mutations that allow it to escape degradation independent of Wnt signaling, AdProt is expressed (*right*).

A ΔAdPolΔAdProt-adenovirus was produced that expressed tBcat from a CMV promoter, tBcat.mCherry (**Table A.5**). After initially passaging the virus once on mutator cells, the virus was then split across eight replicates and passaged 6 times in parallel on selector cells by transferring 5 mL from one plate to the next every 5–10 days (**Figure A.2a**). Each passage would generally start with an MOI of 0.1–0.5, and would end when 50–100% of the cells had died (**Figure A.2b**). Interestingly, while monitoring passage 7 via phase-contrast and fluorescence microscopy, the virus seemed to replicate much more quickly in most of the replicates than in previous passages, leading to earlier CPE (**Figure A.2c**). We subsequently increased the selection stringency by decreasing the volume passaged from 5 mL to 1 mL.

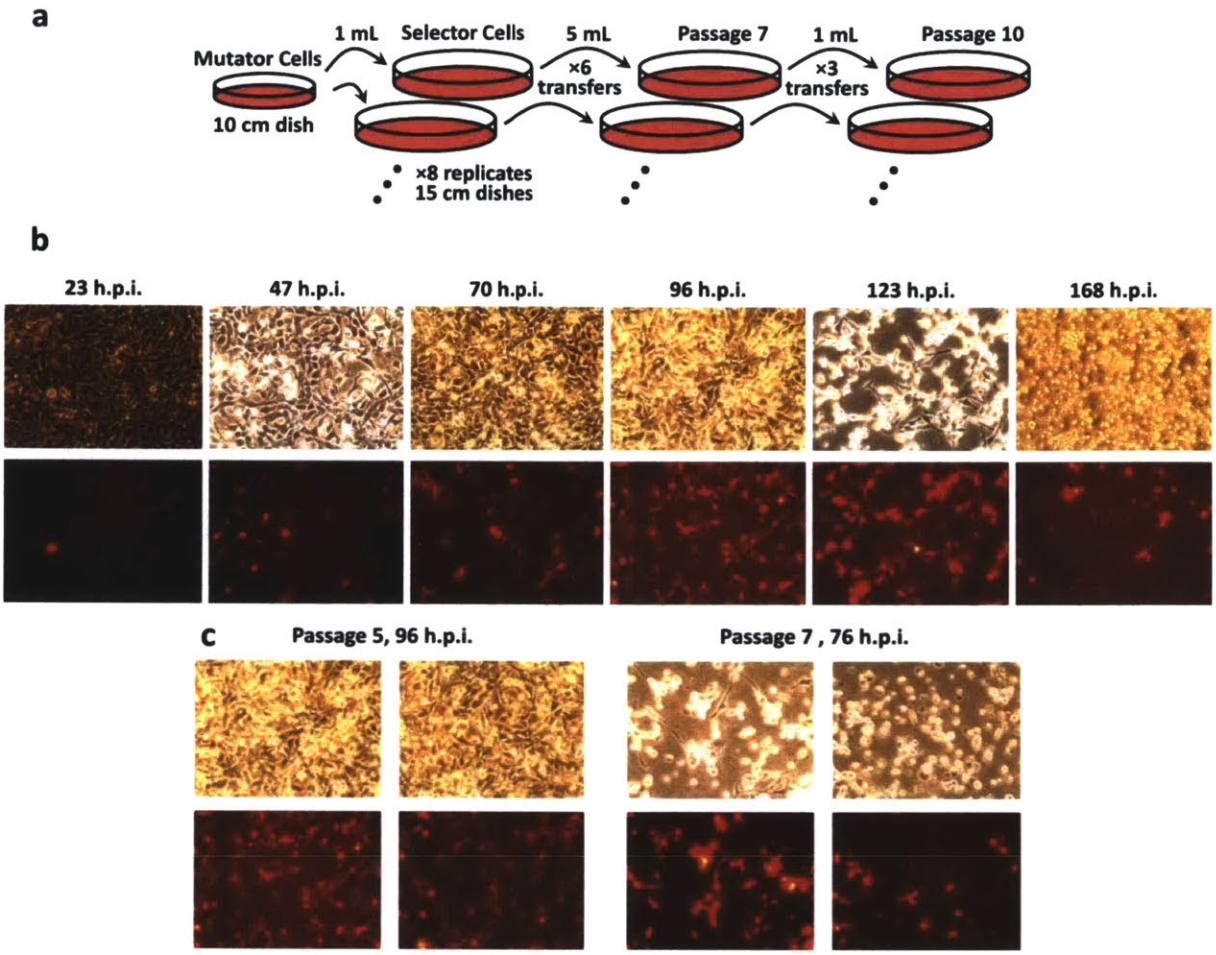

Figure A.2 Passaging tBcat-containing virus on selector cells.
(**a**) Scheme for passaging of tBcat-containing virus. (**b**) Representative phase contrast (*top*) and mCherry fluorescence (*bottom*) images of passage 5 as the infection progressed. Full CPE was observed by 168 hours post-infection. (**c**) Comparison of passage 5 (96 hours post-infection) and passage 7 (76 hours post-infection).

After passage 8, the tBcat gene was PCR amplified from each viral population and Sanger sequenced. Each of the eight populations still had the initial tBcat sequence, with no obvious consensus mutations having fixed in any of the populations. Each replicate was then passaged for two additional passages. During passage 10, replicates 2 and 6 had a much lower MOI than the other replicates, suggesting that these two replicates were impacted by the increased selection stringency of the previous two passage more than the other replicates. Replicates 1, 3, 4, 5, 7, and 8 had very likely adapted somehow, but we had difficulty amplifying the tBcat gene to investigate. The difficulty of amplifying the tBcat gene raised suspicion that a recombination event may have replaced the tBcat gene.

In most adenoviral vectors, including those used here, the E1 region is deleted and generally replaced by an exogenous BOI. Deleting the E1 region allows larger BOIs to be inserted into the genome without exceeding the ~38 kb capacity of the adenoviral capsid[23] and prevents the replication of the engineered adenoviral vector in normal human cells.[24]

The E1 genes are responsible for causing quiescent cells to enter S phase and begin synthesizing DNA, among other roles, which is a required step in the adenoviral life cycle and can also be used to immortalize certain human cell lines.[25] HEK293 cells and their derivatives, such as all cell lines used in this work, were in fact generated in the 1970s when human embryonic kidney cells (HEK) were immortalized by transfection with sheared adenoviral genome fragments and subsequent integration of the left-hand end of the adenoviral genome into the HEK genome.[26] Sequencing in 1997 later confirmed that a fragment of the adenoviral genome consisting of nucleotides 1 through 4344, encompassing the E1 genes, had integrated into the HEK293 genome.[27]

Because HEK293 cells express the exogenous E1 genes, they can support the replication of E1-deleted adenoviruses, and are the main cell line used for this purpose by us and others. However, homologous recombination between the HEK293 genome-integrated copy of the E1 genes and E1-deleted vectors has been observed (**Figure A.3a**),[24] with the resulting E1-containing recombinant viruses potentially have an advantage over E1-lacking viruses.[28]

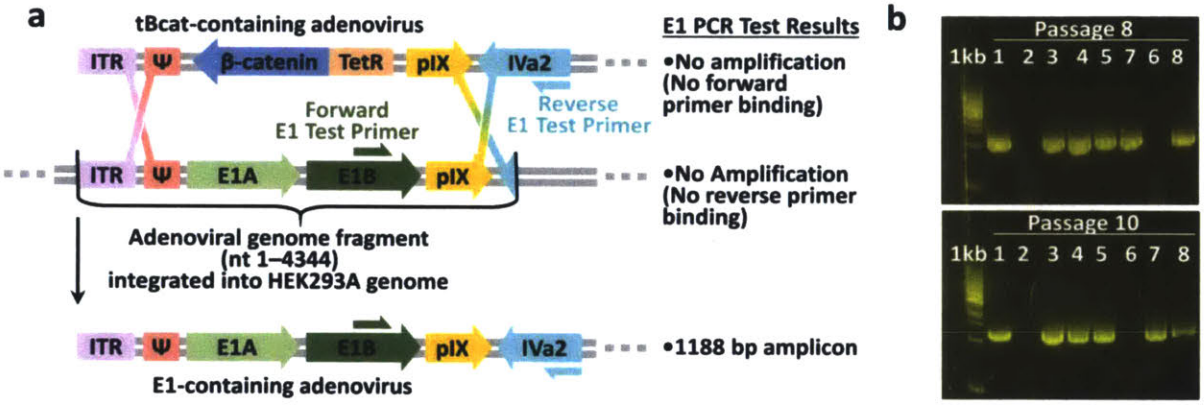

Figure A.3 Replacement of tBcat with E1 genes via recombination.
(**a**) HEK293A cells contain a fragment of the adenoviral genome that can undergo homologous recombination with E1-deleted adenoviral vectors and yield E1-containing adenoviruses (**b**) Using a pair of primers that will only produce an amplicon if crossover has occurred between the HEK293A genome and the adenoviral vector, it is apparent E1-containing virus is present replicates 1, 3, 4, 5, 7, and 8. MW markers are indicated by "1 kb" (New England Biolabs 1 kb ladder B7025). Note that replicates 7 and 6 were accidentally switched in passage 8 when loading the gel, which is reflected in the numbering.

We conducted a PCR test on the passage 8 and passage 10 populations to determine whether there were any recombinant viruses in any of the replicates that had reacquired the E1 region. Surprisingly, all of the replicates except 2 and 6 were positive of the presence of E1-containing virus in both passage 8 and passage 10 (**Figure A.3b**), which suggests that the adaptation in replicates 1, 3, 4, 5, 7, and 8 that lead to a higher rate of propagation was the replacement of tBcat with the E1 genes via recombination. It is also possible that the adenoviruses in replicates 1, 3, 4, 5, 7, and 8 also reacquired the AdProt genes to cheat the selection couple, however this did not appear to be the case when tested via PCR (**Figure A.4**). Taken together, this evidence suggests that having the E1 genes in the adenoviral genome rather than the HEK293 genome confers a greater selective advantage than the tBcat gene alone, since the replicates containing viruses that replaced the tBcat gene with the E1 genes thrived more than those that had not. The

dominance of E1-containing viruses lacking tBcat can be explained by multiple potential scenarios, either:

1) the E1-containing viruses acquired an additional adaptation to either induce the AdProt gene without tBcat or depend less on AdProt,

2) the basal AdProt expression without activation by tBcat is sufficient enough for viruses lacking tBcat to replicate robustly, or

3) there is a mixture of wild-type tBcat.mCherry and E1-containing viruses cooperating, with the E1-containing viruses relying on the un-evolved tBcat viruses inducing high AdProt levels, and the tBcat viruses benefiting from a higher expression level of E1 genes via the E1-containing viruses.

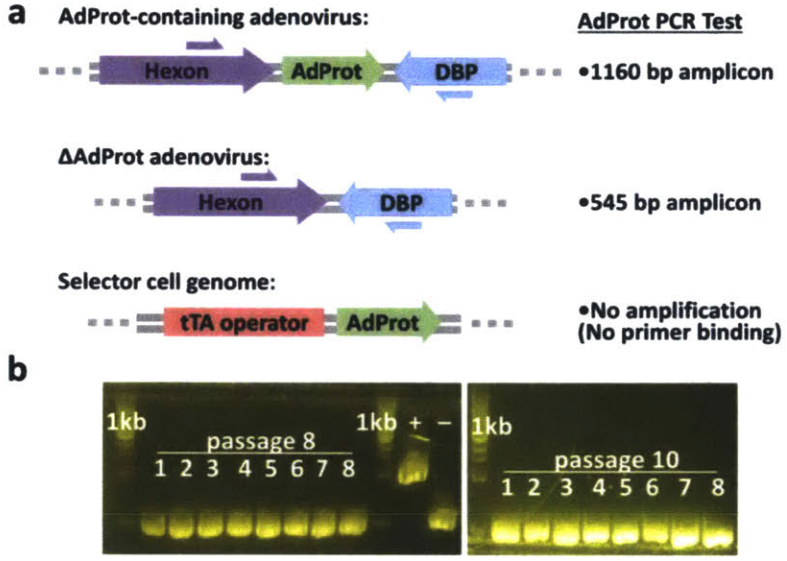

Figure A.4 E1-containing viruses did not reacquire AdProt.
(**a**) Overview of the AdProt PCR test. Primers were designed such that AdProt-containing and ΔAdProt adenoviruses yielded amplicons of significantly different sizes. Primers were also designed to avoid producing amplicons from potential contamination of selector cell genomic DNA. (**b**) The AdProt PCR test suggests that all eight replicates in passage 8 and passage 10 contain ΔAdProt adenovirus. An AdProt-containing adenoviral plasmid was used as a positive control (+) and a ΔAdProt adenoviral plasmid was used as a negative control (–). MW markers are indicated by "1 kb" (New England Biolabs 1 kb ladder B7025).

We believe that scenario 1 is unlikely because AdProt plays a central role in adenoviral replication[29,30] and because it seems unlikely that the adenoviral genome could evolve an alternate mechanism to induce the tTA operators. Scenario 2 provides the simplest explanation assuming that wild-type tBcat is unnecessary and does not provide a greater advantage than the E1 genes. The fact that the same cell line was used to evolve tTA variants that clearly conferred a selective advantage suggests that perhaps there are variants of tBcat that could potentially confer a greater advantage than the E1 genes, but that had failed to appear before E1-containing virus outcompeted and eliminated tBcat.mCherry viruses. Another possibility with scenario 2 is that the tBcat fusion is non-functional, unlike the previously reported β-catenin–Lef1 fusion,[22] and

viruses that replace it with the beneficial E1 genes suffer no potential disadvantage. Scenario 3 is possible since there were tBcat.mCherry and E1-containing adenoviruses clearly co-existed at passage 8 in replicates 1, 3, 4, 5, 7, and 8, and it is uncertain if tBcat.mCherry virus had completely vanished by passage 10.

If scenario 3 is true, it suggests that the parental variant of tBcat persisted because it already had maximal AdProt expression activity and there was no selection pressure to evolve. To test this possibility, we infected selector cells in parallel with adenovirus carrying the parent wild-type tBcat or an oncogenic variant of β-catenin that is known to evade degradation, S45F,[31] and monitored the infection by microscopy (**Figure A.5**). Both the wild-type tBcat variant and the S45F tBcat variant appeared to propagate at similar speeds, suggesting that the S45F mutation did not confer a significant advantage to the adenovirus carrying that variant. The lack of a difference between wild-type and S45F tBcat suggests that either the tBcat fusion protein is non-functional with or without the S45F mutation (supporting scenario 2 in which basal AdProt expression is sufficient for viral propagation), or the tBcat fusion is overexpressed to such a high level that it overwhelms the degradation machinery without even requiring the acquisition of mutations such as S45F (supporting scenario 3 in which wild-type tBcat.mCherry virus and E1-containing virus cooperate).

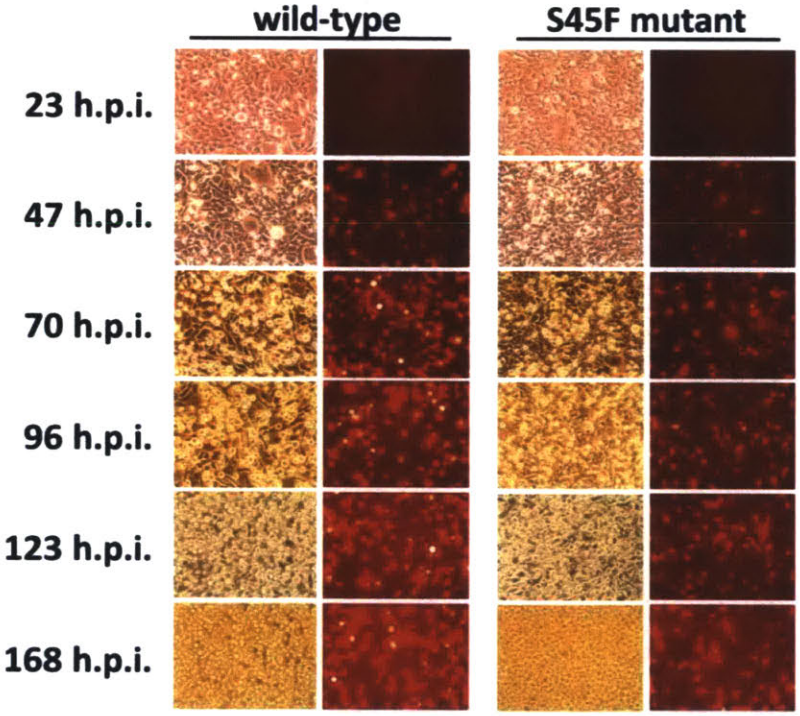

Figure A.5 Competition between wild-type tBcat and the oncogenic S45F variant. Representative phase contrast (*first and third column*) and mCherry fluorescence (*second and fourth column*) images of selector cells infected with adenovirus carrying wild-type tBcat or the S45F variant. Full CPE was observed by 168 hours post-infection.

Since adenoviruses containing wild-type β-catenin have previously been employed to increase β-catenin levels to the point of signal activation,[32] it is conceivable that our viruses also express wild-type tBcat to excessive levels that allow for downstream signal activation without upstream Wnt signaling or oncogenic mutations. Expression from the

strong CMV promoter in our adenoviral constructs likely increases tBcat levels to a level that overwhelms the capacity of the cell to degrade even the wild-type variant. To test this scenario, we redid the experiment using viruses that express the tBcat gene using the weak minCMV promoter and increased the selection pressure by passaging only 250 µL of media from plate to plate in eight replicates. We did not include an initial passage on mutator cells as we did in the first experiment, since any E1-containing viruses that possibly show up in this common passage will contaminate all eight subsequent replicates. The starting MOI of each passage was generally below 0.1 and each passage took longer to reach 50–100% CPE, generally 9–16 days (**Figure A.6a**).

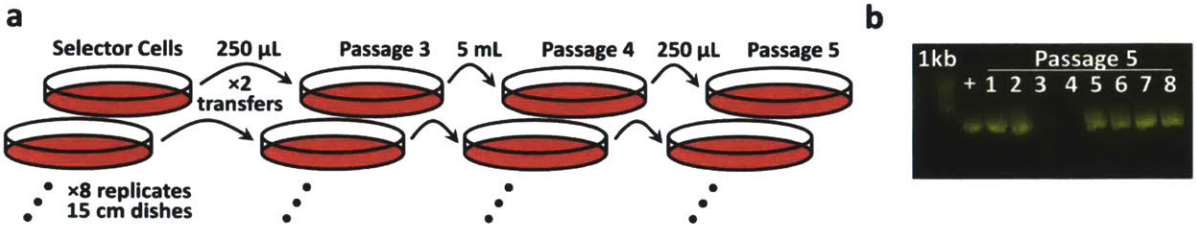

Figure A.6 Passaging minCMV tBcat-containing virus yields E1-containing virus.
(**a**) Scheme for passaging of minCMV tBcat-containing virus. (**b**) An E1 PCR test suggests the presence of E1-containing virus in six out of the eight replicates at passage 5. An E1-containing viral DNA sample from passage 10 of the previous tBcat passaging experiment was used as a positive control (+). MW markers are indicated by "1 kb" (New England Biolabs 1 kb ladder B7025).

Very little virus was produced after passage 3, and we found that we needed to transfer 5 mL onto passage 4 rather than 250 µL. Despite these apparently very stringent conditions, the virus harvested from passage 5 contained E1 genes in six out of the eight replicates, with the other two replicates lacking the E1 genes and having visibly reduced propagation (**Figure A.6b**). Furthermore, consensus Sanger sequencing of the β-catenin N-terminal tail of the tBcat gene in seven of the replicates revealed the wild-type sequence with no obvious fixed mutations, and we were unable to amplify the tBcat gene from the eighth replicate. Given that no new variants of tBcat fixed in the population and that the minCMV promoter is roughly 5-fold weaker than the CMV promoter in HEK293T cells,[33] scenario 3 seems less likely since the minCMV promoter should ultimately be less capable of inducing high AdProt levels and should have been forced to evolve in order to propagate and support E1-containing virus. Scenario 2 seems more likely, where E1-containing adenovirus is not relying on tBcat co-infection and induction at all, and rather the basal levels of AdProt are sufficient for E1-containing virus to replicate alone.

If basal levels of AdProt are sufficient to support the replication of tBcat-lacking virus and the S45F mutation does not confer a selective advantage to tBcat containing cells, the tBcat fusion is likely non-functional. The successful tTA evolution proves that induction of AdProt above the basal level of the current selector cell line can indeed confer a selective advantage, but tBcat apparently failed to increase AdProt expression with or without a known oncogenic mutation. In light of the apparent non-viability of a TetR-tBcat fusion and the frequent appearance of unwanted E1-containing virus, we decided to redesign both the adenoviruses and the cell lines. We created a panel of viruses containing unfused wild-type β-catenin with either the CMV promoter, minCMV promoter,[33] or JeT promoter,[34] which is 4-fold weaker than the CMV promoter. While the

original tBcat fusion contained an HA epitope tag between the TetR and β-catenin domains for quantifying tBcat levels, the panel includes β-catenin with and without N-terminal HA-epitope tags in the event that the tag interferes with β-catenin function. We have also included an oncogenic variant of β-catenin, S45F,[31] in the panel to serve as a positive control in future experiments (**Table A.1**).

Table A.1 Panel of unfused β-catenin viruses.

Promoter	N-terminal epitope	β-catenin
CMV	None	wild-type
CMV	None	S45F
CMV	HA	wild-type
CMV	HA	S45F
minCMV	None	wild-type
minCMV	None	S45F
minCMV	HA	wild-type
minCMV	HA	S45F
JeT	None	wild-type
JeT	None	S45F
JeT	HA	wild-type
JeT	HA	S45F

We also constructed a cell line that expresses AdProt from a synthetic TCF/Lef1 responsive promoter.[35] This circuit will relies on the endogenous β-catenin signaling transcription co-factors already present in human cells, thus allowing for the evolution of β-catenin in its natural context (**Figure A.7a**). Furthermore, this new cell line was derived from GH329 cells which contain the E1 genes but, unlike HEK293 cells, only share homology on one side of the adenoviral vectors and should be incapable of producing E1-containing virus via double-crossover homologous recombination (**Figure A.7b**).[36]

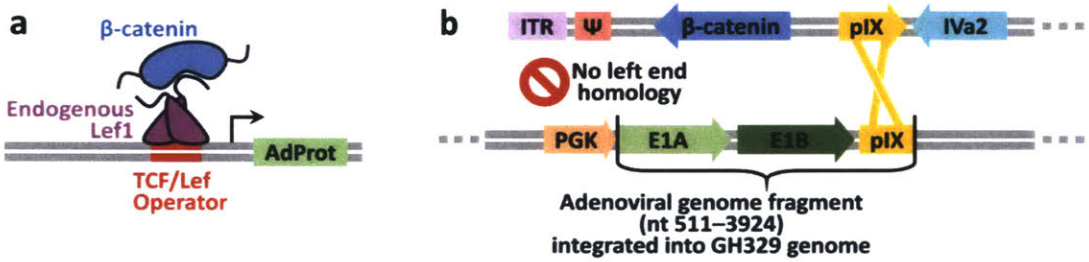

Figure A.7 Selection circuit utilizing unfused β-catenin and the GH329 cell line. (**a**) If β-catenin acquires a mutation that allows it to escape degradation and accumulate, it will induce AdProt expression from a TCF/Lef operator using endogenous transcription co-factors. (**b**) The GH329 cell line expresses the E1 genes from the PGK promoter and does not share homology on the left end of E1-deleted adenoviral vectors, thus preventing the production of E1-containing virus via double-crossover homologous recombination.

Finally, it is worth noting that, to date, the above evolutions have only been attempted with selection circuits that are stably integrated into the host cell genome. While stable integration allows for reproducibility from passage to passage and obviates the need to transfect cells with the selection circuit before each passage, it is possible that the low number of copies of the selection circuit in stable cell lines results in circuits that are not highly dosable and that are easily saturated. It may be worth attempting future

evolutions with transiently transfected cells to assess whether there is a difference in dosability or the level of enrichment.

A.4 Discussion

We attempted to evolve β-catenin, a transcription factor that functions as a central component in a complex human cellular signaling pathway that is not easily recapitulated in *E. coli* or yeast. We elected to use our existing selector cells since they already enabled a successful evolution experiment. In an attempt to couple the transcriptional activation activity of β-catenin to the tTA operator in our previously reported select cells,[37] we fused β-catenin to TetR, which binds tTA. β-catenin has been shown previously to tolerate fusion to the Lef1 transcription factor,[22] but it is unknown if the TetR–β-catenin fusion is actually able to induce transcription from the tTA operator. After 8 passages on selector cells, no mutations had fixed in the tBcat.mCherry viruses as determined by Sanger sequencing. Furthermore, six out of the eight evolution replicates had acquired E1-containing adenoviruses through homologous recombination, and the viruses in these replicates clearly propagated more rapidly than those in replicates without E1-containing adenovirus. Furthermore, a side-by-side comparison of adenoviruses containing wild-type tBcat and a version of tBcat with a mutation known to confer high activity to β-catenin showed no clear difference in the rate of propagation, suggesting that either wild-type tBcat already induces AdProt expression at a maximal rate, or that the tBcat fusion protein is non-function with or without the S45F mutation.

Repeating the evolution experiment with a weaker promoter driving tBcat expression yielded similar results, with no fixed mutations after 5 passages and the appearance of E1-containing virus in six out of eight replicates. The fact that an even weaker promoter failed to increase selection pressure and result in tBcat evolution suggests that either starting tBcat activity is still too high, or more likely the tBcat fusion protein is non-functional and basal AdProt levels are high enough for ΔAdProt virus to propagate with or without tBcat. It is unclear whether E1-containing virus is deleterious to the success of the evolution experiment, such as by outcompeting potential variants that could have arisen, or whether the E1-containing virus is generally a harmless contaminant that only takes over in the absence of significant selection pressure for the presence of the BOI. The successful evolution of the tTA transcription factor seems to suggest the latter. Given the unknown impact of E1-containing virus and the unknown viability of a TetR–β-catenin fusion, we constructed a new selection circuit that expresses AdProt from an operator that is endogenously activated by unfused, natural β-catenin and integrated it into a cell line that is incapable of generating E1-containing virus,[36] unlike the HEK293-derived cell lines we initially used. Efforts to validate and utilize this new system are ongoing.

While we have not yet demonstrated successful evolution of β-catenin, our initial attempts were quite informative for improving the system. We learned that E1-containing virus propagates much more rapidly than E1-deleted virus, even though the E1 genes are already constitutively expressed in all of our HEK293 derived cell lines. To prevent the E1 crossover event from replacing the BOI in future evolutions, we can use special cell lines that express the E1 genes, but that do not share homology with our adenoviral vectors. A potentially more attractive alternative is restore the E1 genes in our adenoviral vectors and place the BOI gene in the E4 region instead of the E1 region. This arrangement avoids the situation where the E1 regions replace the BOI, and can potentially accelerate our evolution experiment by conferring the apparent advantage of natural E1 expression from the adenoviral genome. This architecture has the added benefit of allowing our

adenoviral vectors to propagate in cells that do not already trans-complement the E1 genes, thus expanded the scope of our platform beyond cell lines derived from HEK293 or similar E1-expressing lines. One disadvantage of placing the E1 genes back into the adenoviral genome is that it decreases the theoretical available capacity for inserted BOIs and fluorescent markers from ~7 kb to ~5 kb. However, the current E3 deletion in most adenoviral vectors, including the ones used here, is only partial and was performed based on the availability of two *XbaI* sites in the E3 region.[38] It has been shown that the entire E3 region can be deleted,[39] so we can regain an additional 1.2 kb of BOI capacity by deleting the residual portion of the E3 region in our vectors (**Table A.2**).

Table A.2 Effect of proposed modifications on BOI packaging capacity

Genotype*	Genome size†	BOI Packaging Capacity	Notes
Maximum capsid limit	~37,700 bp	–	~105% wild-type[23]
Wild-type adenovirus type-5	35,934 bp	~1,700 bp	
ΔE1 ΔE3 ΔAdProt ΔAdPol	29,687 bp	~8,000 bp	Used in this work
E1⁺ ΔE3 ΔAdProt ΔAdPol	32,866 bp	~4,800 bp	Proposed E1 reinsertion
ΔE1 ΔE3-*dl7001* ΔAdProt ΔAdPol	28,501 bp	~9,200 bp	Proposed complete E3 deletion
E1⁺ ΔE3-*dl7001* ΔAdProt ΔAdPol	31,680 bp	~6,000 bp	Proposed E1 reinsertion and complete E3 deletion

* ΔE3 denotes the conventional partial E3 deletion, ΔE3-dl7001 denotes complete E3 deletion
†excluding exogenous DNA such as BOIs, fluorescent proteins, promoters, and polyadenylation signals

We have also learned from our initial β-catenin evolution attempts that the basal "leaky" AdProt expression activity from our selector cells can be high enough to allow for the propagation of viruses that lack a BOI that binds the tTA operator. While some basal leaky propagation is not necessarily bad, and could even be important for allowing the evolution of biomolecules with very weak starting activity, it could also theoretically decrease selection pressure so that only large, sudden improvements in activity are enriched, and incremental small improvements in activity are not advantageous. One strategy to potentially mitigate leaky AdProt expression is to apply adenoviral protease inhibitor when we repeat the β-catenin evolution experiments. Another possibility is that the stably integrated circuit is not dosable enough to generate selection pressure for new variants, and becomes saturated easily. This problem could potentially be resolved by transiently transfecting the selection circuits into the host cell instead. It could also be worth exploring a selection gene that is not enzymatic, unlike AdProt, and is required at higher copy numbers than AdProt, such that even significantly leaky expression is insufficient to propagate viruses lacking the desired BOI activity. While AdProt is the first selection gene we have tested and has proved sufficient to evolve tTA, there are other genes that have been successfully trans-complemented in the literature that may have even more beneficial properties for our platform, such as fiber,[40-42] pTP,[43] DBP,[44,45] AdPol,[46] and 100K.[47]

Along with the potential improvements gleaned from our attempted evolution of β-catenin, we envision additional modifications that would further enhance this platform's practicability and applicability. The adenoviral death protein (ADP), which is normally lost

upon E3 deletion and thus absent in most adenoviral vectors, is important for cell lysis and viral release.[48] It has been shown that reintroducing ADP into an E3-deleted vector greatly accelerates adenoviral vector propagation,[49] and would likely allow us to shorten the time it takes to produce virus or to complete a round of evolution.

A.5 Conclusion

While initial attempts to evolve β-catenin did not yield any interesting new variants, these experiments yielded important insights on how to further improve our platform. Moving BOI-encoding genes to the E4 region can prevent the loss of the BOI gene due to E1 recombination and/or E1 recombination can be prevented with E1-complementing cell lines that lack homology with the adenoviral vector. We also found that restoring the E1 genes in our adenoviral vectors significantly increase the rate at which the virus propagates, potentially providing a means for the accelerations of directed evolution experiments using our platform. With a few modifications, we anticipate successfully exploring the drug-resistance mutational trajectories available to β-catenin under different cellular conditions.

A.6 Acknowledgements

This work was supported by an NIH Director's New Innovator Award (1DP2GM119162), MISTI Global Seeds Fund, and MIT (all to M.D.S.) and by the European Union through the 6th Framework Program GIANT (contract no. 512087; to Robert C. Hoeben). L.J.P. and S.J.H. were supported by National Science Foundation Graduate Research Fellowships under Grant No. 1122374. This work was also supported in part by the NIH/NIEHS under award P30-ES002109 and by Cancer Center Support (core) Grant P30-CA14051 from the NIH/NCI.

A.8 Materials and Methods

Cloning methods: All PCR reactions for cloning and assembling recombineering targeting cassettes were performed using Q5 High Fidelity DNA Polymerase (New England BioLabs). Restriction cloning was performed using restriction endonucleases and Quick Ligase from New England BioLabs. Adenoviral constructs were engineered using ccdB recombineering, as previously described[50] and further optimized by us.[51] Primers were obtained from Life Technologies and Sigma-Aldrich (**Table A.3**). Plasmid sequences can be obtained from GenBank using the accession numbers provided in **Table A.4**. The pDONR223_CTNNB1_WT and pDONR223_CTNNB1_p.S45F plasmid were gifts from Jesse Boehm & William Hahn & David Root (Addgene plasmid #81734 and #82821, respectively).[52] The M50 Super 8x TOPFlash plasmid was a gift from Randall Moon (Addgene plasmid #12456).[53]

pLVX.TCF.AdProt vector: A 233 bp fragment containing an array of TCF/Lef binding sites was amplified from M50 Super 8x TOPFlash[53] using the primers ClaI.TCF.Forward and BamHI.TCF.Reverse (**Table A.3**) and ligated into the pLVX.TRE3G.AdProt vector (**Table A.4**) using ClaI and BamHI to create the pLVX.TCF.AdProt vector.

Table A.3 Primers used for cloning and recombineering.

Name	Sequence
ClaI.TCF.Forward	5′-aaaaaaatcgataggtaccgagctcttac-3′
BamHI.TCF.Reverse	5′-aaaaaaggatccgtaccggaatgccaagctgg-3′
TetR.kanccdB.Forward	5′-tggaactaatcatatgtggcctggagaaacagctaaagtgcgaaagcggcccgctcatta ggcgggc-3′
TetR.kanccdB.Reverse	5′-cgcgaacaaatgtggtatggctgattatgatcctctagagataattctagccctcatcagtgcc aacatagtaag -3′
TetR.HA.Bcat.Forward	5′-actaatcatatgtggcctggagaaacagctaaagtgcgaaagcggctacccatacgacgt accagattacgctatggctactcaagctgatttgatggag-3′
Bcat.Reverse	5′-cgcgaacaaatgtggtatggctgattatgatcctctagagataattctagttacaggtcagtat caaaccaggcc-3′
minCMV.Forward	5′-agctctgcttatatagacctcccaccgtacacgcctaccaacgcgtatatctgtgaattaaaa aagaattcaaaaaaaagcccgc -3′
minCMV.Reverse	5′-cccacccttaagccacgcccacagatatacgcgttggtaggcgtgtacgggaattaaaaa agaattcaaaaaaaagcccgc-3′
AmilCP.Kan.Reverse	5′-cgccgtcgggcatgc-3′
Kan.AmilCP.Forward	5′-gctcgacgttgtcactgaagc-3′
HA.Bcat.Forward	5′-tacgactcactatagggagacccaagctggctagttaagctatcaatgtacccatacgatgtt ccagattacgctgctactcaagctgatttgatgg-3′
Bcat.Forward	5′-attaatacgactcactatagggagacccaagctggctagttaagctatcaatggctactcaag ctgatttgatgg-3′
JeT.kanccdB.Forward	5′-ccttatatattctttcccacccttaagccacgcccacagatatacgcgttcctgttgacaattaat catcggca-3′
JeT.kanccdB.Reverse	5′-gtattaatttcgataagccagtaagcagtgggttctctagttagccagagccgctcattaggc gggc-3′
JeT.Forward	5′-ccttatatattctttcccacccttaagccacgcccacagatatacgcgttgaattcgggcggagt tagggc-3′
JeT.Reverse	5′-gtattaatttcgataagccagtaagcagtgggttctctagttagccagagtgtcaagtgacgat cacagggatcc-3′
JeT gene block	5′-gcagaacgcccggcacgtgctctgccagttgttcaatggcctgattcgagcaattgctcgag

	gggcggagttagggcggagccaatcagcgtgcgccgttccgaaagttgccttttatggctgggcg gagaatgggcggtgaacgccgatgattatataaggacgcgccgggtgtggcacagctagttcc gtcgcagccgggatttgggtcgcggttcttgtttgtggatccggcatcgatgctagcaccatgccaa aaaagaagagaaaggtaggaacgcgtaccatgtacccatacgacgtaccagattacgctatgt cttctgaaaccggtccggttgcggttgacccgaccctgcgtcgtcg-3′
E1Test.Forward	5′-ggaaccagcctgtgatgctg-3′
E1Test.Reverse	5′-aaaatcttggaggtcacaagggc-3′
AdProtTest.Forward	5′-gcttataggcaagaccgcag-3′
AdProtTest.Reverse	5′-cggatggttgtgcctgagtttaag-3′
CMV.seq.Forward	5′-cagatatacgcgttgacattg-3′
HSV.seq.Reverse	5′-gaagccatagagcccac-3′
Bcat.seq.1	5′-agcagctgcatatgtcgcc-3′
Bcat.seq.2	5′- tgcattctgggccatctctgc-3′
Bcat.seq.3	5′-tgccataagctaaaatttgaaggcagtc-3′
Bcat.seq.4	5′-gcatctgtgatggttcagccaaac-3′
Bcat.seq.5	5′-caaagagtagctgcaggggtcc-3′
tBcat.seq.Forward	5′-aaaaaagtcgaccctctagtcagctgacgcg-3′
tBcat.seq.Reverse	5′-aaaaaaggatccgcgaacaaatgtggtatggctg-3′
E1.kanccdB.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccacccttaaccctcatcagtgcca acatagtaag-3′
E1.kanccdB.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaaccgctcattaggcg ggc-3′
E1.CMV.Promoter.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccacccttaagccacgcccacaga tatacgcgttgacattg-3′
E1.bGH.polyA.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaatagaagccataga gcccac-3′

Table A.4 Plasmid sequence accession numbers.

Vector Name	GenBank Accession Number
pLVX.TRE3G.eGFP	MH325104
R6K-kan-ccdB	MH325106
R6KAmilCP-kan-ccdB	MK787296
R6K-kan-ccdBAmilCP	MK787297

Table A.5 Adenoviruses constructed and used in this study.

Name	Modifications relative to wild-type Ad5
tTA_{wt}.mCherry	E1L-tTA ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
tBcat.mCherry	E1L-tBcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
$tBcat_{S45F}$.mCherry	E1L-tBcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
minCMV-tBcat.mCherry	E1L-minCMV-tBcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
minCMV-$tBcat_{S45F}$.mCherry	E1L-minCMV-tBcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
Bcat.mCherry	E1L-Bcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
$Bcat_{S45F}$.mCherry	E1L-Bcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
HA-Bcat.mCherry	E1L-HA-Bcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
HA-$Bcat_{S45F}$.mCherry	E1L-HA-Bcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
minCMV-Bcat.mCherry	E1L-minCMV-Bcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
minCMV-$Bcat_{S45F}$.mCherry	E1L-minCMV-Bcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
minCMV-HA-Bcat.mCherry	E1L-minCMV-HA-Bcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
minCMV-HA-$Bcat_{S45F}$.mCherry	E1L-minCMV-HA-Bcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
JeT-Bcat.mCherry	E1L-JeT-Bcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
JeT-$Bcat_{S45F}$.mCherry	E1L-JeT-Bcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
JeT-HA-Bcat.mCherry	E1L-JeT-HA-Bcat ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
JeT-HA-$Bcat_{S45F}$.mCherry	E1L-JeT-HA-Bcat-S45F ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry

Adenoviral constructs: Recombineering was performed as previously described.[37,51] The following modifications (**Table A.6**) were made using the primers in **Table A.3** to obtain the adenoviruses (**Table A.5**) used in this work. All viruses were derived from tTA_{WT}.mCherry (**Table A.5**):

Table A.6 Modifications made to the adenoviral genome.

Modification	Genotype	*Kan^R+ccdB* cassette primers used with R6K-kan-ccdB template plasmid (unless stated otherwise)	Final targeting cassette oligos or primers and template (if applicable)	Purpose of modification
Fusion of β-catenin to TetR	E1L tBcat or E1L tBcat-S45F	TetR.kanccdB.Forward and TetR.kanccdB.Reverse	TetR.HA.Bcat.Forward and Bcat.Reverse used to amplify from pDONR223_CTNNB1_WT or pDONR223_CTNNB1_p.S45F	To fuse wild-type or S45F β-catenin to the C-terminus of TetR in tTA$_{WT}$.mCherry with an intervening HA epitope tag.
Creation of minCMV via deletion of CMV enhancer	E1L-minCMV-	minCMV.Forward and AmilCP.Kan.Reverse used to amplify from R6KAmilCP-kan-ccdB template plasmid and minCMV.Reverse and Kan.AmilCP.Forward used to amplify from R6K-kan-ccdBAmilCP template plasmid	None, see below for descriptions of one-step DIRex recombineering.	Deletes the enhancer from the CMV promoter to make it weaker.
Insertion of unfused β-catenin	E1L-Bcat	TetR.kanccdB.Forward and TetR.kanccdB.Reverse	HA.Bcat.Forward and Bcat.Reverse used to amplify from pDONR223_CTNNB1_WT or pDONR223_CTNNB1_p.S45F	To insert β-catenin into the E1L region, completely replacing tTA.
Insertion of unfused β-catenin with HA epitope tag.	E1L-HA-Bcat	TetR.kanccdB.Forward and TetR.kanccdB.Reverse	Bcat.Forward and Bcat.Reverse used to amplify from pDONR223_CTNNB1_WT or pDONR223_CTNNB1_p.S45F	To insert β-catenin into the E1L region with an N-terminal HA epitope tag, completely replacing tTA.
Replacement of the CMV promoter with the JeT promoter	E1L-JeT	JeT.kanccdB.Forward and JeT.kanccdB.Reverse	JeT.Forward and Jet.Reverse used to amplify from JeT gene block	To replace the strong CMV promoter with the weak JeT promoter.

Converting the CMV promoter to minCMV using DIRex lambda red recombineering: The enhancer regions of the CMV promoter were deleted using one-step DIRex recombineering,[54] with modifications to the protocol made by us.[51] The minCMV targeting half-cassettes were amplified from R6KAmilCP-kan-ccdB (**Table A.4**) using the primers minCMV.Forward and AmilCP.Kan.Reverse (**Table A.3**) and from R6K-kan-ccdBAmilCP (**Table A.4**) using the primers minCMV.Reverse and Kan.AmilCP.Forward (**Table A.3**). The minCMV half cassetteswere co-electroporated to replace the CMV enhancer 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[54] that was selected for by simply streaking the kanR+ccdB intermediates on LB agar plates with 10 μg/mL chloramphenicol that lacked arabinose. The resulting colonies were screened for deletion of the CMV enhancer, yielding the minCMV promoter.

Cell culture: Performed as previously described.[37]

Continuous tBcat evolution workflow: Before initiating directed evolution, 200 μL of a tBcat.mCherry adenovirus was amplified on a 10 cm dish of confluent mutator cells[37] to create a diverse viral population. After 8 days, cytopathic effect was observed in all cells. This amplified virus was harvested with three freeze/thaw cycles. Eight 15 cm, semi-confluent dishes of selector cells (~1 × 10^7 cells/plate) were infected with 1 mL of the amplified virus. Plates were monitored for CPE every day. Once 50–100% of the cells exhibited CPE, typically every 5–10 days, 5 mL of the media would be transferred to eight fresh, semi-confluent, 15 cm dishes of selector cells and 2 mL of media were saved in Eppendorf tubes and stored at –80 °C for future analysis. Each passage would generally start with an MOI of 0.1–0.5, and would end when 50–100% of the cells had died. After six 5 mL transfers, at passage 7, it was apparent that the viral populations were replicating much more quickly, so only 1 mL of media was transferred in the subsequent 3 passages to increase selection stringency. Pictures were taken with a Nikon Eclipse TE200 microscope.

E1 and AdProt PCR tests: DNA was purified from 200 μL of media from each viral passage using a viral DNA isolation kit (NucleoSpin Virus; Macherey-Nagel). The DNA was eluted with with 30 μL MBG water to generally yield ~150ng/μL. To test for the presence of E1-containing virus, 1 μL of each viral prep was add used as a template for PCR using the E1Test.Forward and E1Test.Reverse primers (**Table A.3**) and the Q5 High Fidelity DNA Polymerase (New England BioLabs) according to manufacturer instructions. The following PCR protocol was executed and the PCR reactions were visualized by gel electrophoresis, with the presence of a band at ~1.1 kb indicating the presence of E1-containing virus:

1. 98 °C	30 sec
2. 98 °C	10 sec
3. 69 °C	30 sec
4. 72 °C	2 min
5. Return to step 2 for 35 cycles	
6. 72 °C	2 min
7. 4 °C	hold

To test for the presence of AdProt-containing virus, 1 μL of each viral prep was add used as a template for PCR using the AdProtTest.Forward and AdProtTest.Reverse primers (**Table A.3**) and the Q5 High Fidelity DNA Polymerase (New England BioLabs) according to manufacturer instructions. The following PCR protocol was executed and the PCR reactions were visualized by gel electrophoresis, with the presence of a band at ~1.1 kb indicating the reacquisition of AdProt and a band at ~0.5 kb indicating AdProt deletion:

1. 98 °C	30 sec
2. 98 °C	10 sec
3. 67 °C	30 sec
4. 72 °C	2 min
5. Return to step 2 for 35 cycles	
6. 72 °C	2 min
7. 4 °C	hold

Consensus Sanger sequencing: DNA was purified from 200 µL of media from each viral passage using a viral DNA isolation kit (NucleoSpin Virus; Macherey-Nagel). The DNA was eluted with with 30 µL MBG. To sequence the tBcat gene during the first passaging experiment, 1 µL of each viral prep was add used as a template for PCR using the CMV.seq.Forward and HSV.seq.Reverse primers (**Table A.3**) and the Q5 High Fidelity DNA Polymerase (New England BioLabs) according to manufacturer instructions. The following PCR protocol was executed and the PCR reactions were checked gel electrophoresis:

1. 98 °C	30 sec
2. 98 °C	10 sec
3. 62 °C	30 sec
4. 72 °C	3 min
5. Return to step 2 for 35 cycles	
6. 72 °C	2 min
7. 4 °C	hold

Each amplicon was purified purified using the E.Z.N.A. cycle pure kit (Omega Bio-tek) according to the manufacturer's instructions. The purified DNA was submitted for Sanger sequencing using the primers CMV.seq.Forward, Bcat.seq.1, Bcat.seq.2, Bcat.seq.3, Bcat.seq.4, and Bcat.seq.5 for full coverage of the tBcat gene (**Table A.3**).

To sequence the tBcat gene during the minCMV-tBcat passaging experiment, 1 µL of each viral prep was add used as a template for PCR using the tBcat.seq.Forward and tBcat.seq.Reverse primers (**Table A.3**) and the Q5 High Fidelity DNA Polymerase (New England BioLabs) according to manufacturer instructions. Different primers were used here because CMV.seq.Forward and HSV.seq.Reverse yielded non-specific amplicons previously. The following PCR protocol was executed and the PCR reactions were checked gel electrophoresis:

1. 98 °C	30 sec
2. 98 °C	10 sec
3. 69 °C	30 sec
4. 72 °C	2 min
5. Return to step 2 for 35 cycles	
6. 72 °C	2 min
7. 4 °C	hold

Each amplicon was purified purified using the E.Z.N.A. cycle pure kit (Omega Bio-tek) according to the manufacturer's instructions. The purified DNA was submitted for Sanger sequencing using the primers Bcat.seq.4 for coverage of the N-terminal tail of β-catenin.

Competition between wild-type tBcat and the oncogenic S45F variant: 100 µL of tBcat.mCherry or $tBcat_{S45F}$.mCherry was added to semi-confluent wells of a 12-well dish of selector cells. Plates were monitored for spreading infection every 24 h, with pictures were taken with a Nikon Eclipse TE200 microscope.

Continuous minCMV tBcat evolution workflow Eight 15 cm, semi-confluent dishes of selector cells (~1 × 10^7 cells/plate) were infected with 250 µL of minCMV-tBcat.mCherry. Plates were monitored for CPE every day. Once 50–100% of the cells exhibited CPE, typically every 9–16 days, 250 µL of the media would be transferred to eight fresh, semi-confluent, 15 cm dishes of selector cells and 2 mL of media were saved in Eppendorf tubes and stored at –80 °C for future analysis. Each passage would generally start with an MOI of less than 0.1, and would end when 50–100% of the cells had died. An exception was made after passage 3, which yielded very little virus, requiring the transfer of 5 mL onto passage 4 rather than 250 µL.

A.9 References

[1] van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A.-P., Tjon-Pon-Fong, M., Moerer, P., van den Born, M., Soete, G., Pals, S., Eilers, M., Medema, R., and Clevers, H. (2002) The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells, *Cell 111*, 241–250.

[2] Moon, R. T., Kohn, A. D., De Ferrari, G. V., and Kaykas, A. (2004) WNT and beta-catenin signalling: diseases and therapies, *Nat. Rev. Genet. 5*, 691–701.

[3] Polakis, P. (2000) Wnt signaling and cancer, *Genes Dev. 14*, 1837–1851.

[4] Li, C.-M., Kim, C. E., Margolin, A. A., Guo, M., Zhu, J., Mason, J. M., Hensle, T. W., Murty, V. V. V. S., Grundy, P. E., Fearon, E. R., D'Agati, V., Licht, J. D., and Tycko, B. (2004) CTNNB1 mutations and overexpression of wnt/β-catenin target genes in WT1-mutant Wilms' tumors, *Am. J. Pathol. 165*, 1943–1953.

[5] Grossmann, T. N., Yeh, J. T., Bowman, B. R., Chu, Q., Moellering, R. E., and Verdine, G. L. (2012) Inhibition of oncogenic Wnt signaling through direct targeting of beta-catenin, *Proc. Natl. Acad. Sci. U.S.A. 109*, 17942–17947.

[6] Sukhdeo, K., Mani, M., Zhang, Y., Dutta, J., Yasui, H., Rooney, M. D., Carrasco, D. E., Zheng, M., He, H., Tai, Y. T., Mitsiades, C., Anderson, K. C., and Carrasco, D. R. (2007) Targeting the beta-catenin/TCF transcriptional complex in the treatment of multiple myeloma, *Proc. Natl. Acad. Sci. U.S.A. 104*, 7516–7521.

[7] Qu, Y., Olsen, J. R., Yuan, X., Cheng, P. F., Levesque, M. P., Brokstad, K. A., Hoffman, P. S., Oyan, A. M., Zhang, W., Kalland, K. H., and Ke, X. (2018) Small molecule promotes beta-catenin citrullination and inhibits Wnt signaling in cancer, *Nat. Chem. Biol. 14*, 94–101.

[8] Gandhirajan, R. K., Staib, P. A., Minke, K., Gehrke, I., Plickert, G., Schlösser, A., Schmitt, E. K., Hallek, M., and Kreuzer, K.-A. (2010) Small molecule inhibitors of Wnt/β-catenin/Lef-1 signaling induces apoptosis in chronic lymphocytic leukemia cells *in vitro* and *in vivo*, *Neoplasia 12*, 326–335.

[9] Hallett, R. M., Kondratyev, M. K., Giacomelli, A. O., Nixon, A. M., Girgis-Gabardo, A., Ilieva, D., and Hassell, J. A. (2012) Small molecule antagonists of the Wnt/beta-catenin signaling pathway target breast tumor-initiating cells in a Her2/Neu mouse model of breast cancer, *PLOS ONE 7*, e33976.

[10] Krishnamurthy, N., and Kurzrock, R. (2018) Targeting the Wnt/beta-catenin pathway in cancer: update on effectors and inhibitors, *Cancer Treat. Rev. 62*, 50–60.

[11] Blair, W. S., Pickford, C., Irving, S. L., Brown, D. G., Anderson, M., Bazin, R., Cao, J. A., Ciaramella, G., Isaacson, J., Jackson, L., Hunt, R., Kjerrstrom, A., Nieman, J. A., Patick, A. K., Perros, M., Scott, A. D., Whitby, K., Wu, H., and Butler, S. L. (2010) HIV capsid is a tractable target for small molecule therapeutic intervention, *PLOS Pathog. 6*, e1001220.

[12] Yin, Y. B., Wang, Y. J., Dang, W., Xu, L., Su, J. H., Zhou, X. Y., Wang, W. S., Felczak, K., van der Laan, L. J. W., Pankiewicz, K. W., van der Eijk, A. A., Bijvelds, M., Sprengers, D., de Jonge, H., Koopmans, M. P. G., Metselaar, H. J., Peppelenbosch, M. P., and Pan, Q. W. (2016) Mycophenolic acid potently inhibits rotavirus infection with a high barrier to resistance development, *Antiviral Res. 133*, 41–49.

[13] Musharrafieh, R., Ma, C. L., and Wang, J. (2018) Profiling the *in vitro* drug-resistance mechanism of influenza A viruses towards the AM2-S31N proton channel blockers, *Antiviral Res. 153*, 10–22.
[14] Phillips, A. M., Gonzalez, L. O., Nekongo, E. E., Ponomarenko, A. I., McHugh, S. M., Butty, V. L., Levine, S. S., Lin, Y. S., Mirny, L. A., and Shoulders, M. D. (2017) Host proteostasis modulates influenza evolution, *eLife 6*, e28652.
[15] Phillips, A. M., Ponomarenko, A. I., Chen, K., Ashenberg, O., Miao, J., McHugh, S. M., Butty, V. L., Whittaker, C. A., Moore, C. L., Bloom, J. D., Lin, Y. S., and Shoulders, M. D. (2018) Destabilized adaptive influenza variants critical for innate immune system escape are potentiated by host chaperones, *PLOS Biol. 16*, e3000008.
[16] Phillips, A. M., Doud, M. B., Gonzalez, L. O., Butty, V. L., Lin, Y. S., Bloom, J. D., and Shoulders, M. D. (2018) Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin, *eLife 7*, e38795.
[17] Geller, R., Pechmann, S., Acevedo, A., Andino, R., and Frydman, J. (2018) Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation, *Nat. Commun. 9*, 1781.
[18] Cowen, L. E., and Lindquist, S. (2005) Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi, *Science 309*, 2185–2189.
[19] Rutherford, S. L., and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution, *Nature 396*, 336–342.
[20] Powers, E. T., and Balch, W. E. (2013) Diversity in the origins of proteostasis networks — a driver for protein function in evolution, *Nat. Rev. Mol. Cell Biol. 14*, 237–248.
[21] de la Roche, M., Rutherford, T. J., Gupta, D., Veprintsev, D. B., Saxty, B., Freund, S. M., and Bienz, M. (2012) An intrinsically labile alpha-helix abutting the BCL9-binding site of beta-catenin is required for its inhibition by carnosic acid, *Nat. Commun. 3*, 680.
[22] Vleminckx, K., Kemler, R., and Hecht, A. (1999) The C-terminal transactivation domain of β-catenin is necessary and sufficient for signaling by the LEF-1/β-catenin complex in Xenopus laevis, *Mech. Dev. 81*, 65–74.
[23] Bett, A. J., Prevec, L., and Graham, F. L. (1993) Packaging capacity and stability of human adenovirus type 5 vectors, *J. Virol. 67*, 5911–5921.
[24] Kovesdi, I., and Hedley, S. J. (2010) Adenoviral producer cells, *Viruses 2*, 1681–1703.
[25] Endter, C., and Dobner, T. (2004) Cell transformation by human adenoviruses, In *Adenoviruses: Model and Vectors in Virus-Host Interactions: Immune System, Oncogenesis, Gene Therapy* (Doerfler, W., and Böhm, P., Eds.), pp 163–214, Springer Berlin Heidelberg, Berlin, Heidelberg.
[26] Graham, F. L., Smiley, J., Russell, W. C., and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5, *J. Gen. Virol. 36*, 59–74.
[27] Louis, N., Evelegh, C., and Graham, F. L. (1997) Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line, *Virology 233*, 423–429.

[28] Lochmuller, H., Jani, A., Huard, J., Prescott, S., Simoneau, M., Massie, B., Karpati, G., and Acsadi, G. (1994) Emergence of early region 1-containing replication-competent adenovirus in stocks of replication-defective adenovirus recombinants (ΔE1 + ΔE3) during multiple passages in 293 cells, *Hum. Gene Ther. 5*, 1485–1491.

[29] Greber, U. F., Webster, P., Weber, J., and Helenius, A. (1996) The role of the adenovirus protease in virus entry into cells, *EMBO J. 15*, 1766–1777.

[30] Webster, A., Leith, I. R., and Hay, R. T. (1994) Activation of adenovirus-coded protease and processing of preterminal protein, *J. Virol. 68*, 7292–7300.

[31] Hamada, S., Futamura, N., Ikuta, K., Urakawa, H., Kozawa, E., Ishiguro, N., and Nishida, Y. (2014) CTNNB1 S45F mutation predicts poor efficacy of meloxicam treatment for desmoid tumors: a pilot study, *PLoS One 9*, e96391.

[32] Kim, S. Y., Lee, J. H., Sohn, K. C., Im, M., Lee, Y., Seo, Y. J., Lee, J. H., and Kim, C. D. (2018) Beta-catenin regulates the expression of cAMP response element-binding protein 1 in squamous cell carcinoma cells, *Ann. Dermatol. 30*, 119–122.

[33] Ede, C., Chen, X., Lin, M. Y., and Chen, Y. Y. (2016) Quantitative analyses of core promoters enable precise engineering of regulated gene expression in mammalian cells, *ACS Synth. Biol. 5*, 395–404.

[34] Tornøe, J., Kusk, P., Johansen, T. E., and Jensen, P. R. (2002) Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites, *Gene 297*, 21–32.

[35] Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B., and Clevers, H. (1997) Constitutive transcriptional activation by a β-catenin-Tcf complex in APC-/- colon carcinoma, *Science 275*, 1784–1787.

[36] Gao, G. P., Engdahl, R. K., and Wilson, J. M. (2000) A cell line for high-yield production of E1-deleted adenovirus vectors without the emergence of replication-competent virus, *Hum. Gene Ther. 11*, 213–219.

[37] Berman, C., Papa, L. J., III, Hendel, S. J., Moore, C. L., Suen, P. H., Weickhardt, A. F., Doan, N. D., Kumar, C. M., Uil, T. G., Butty, V. L., Hoeben, R. C., and Shoulders, M. D. (2018) An adaptable platform for directed evolution in human cells, *J. Am. Chem. Soc. 140*, 18093–18103.

[38] Graham, F. L., and Prevec, L. (1991) Manipulation of adenovirus vectors, *Methods Mol. Biol. 7*, 109–128.

[39] Ranheim, T. S., Shisler, J., Horton, T. M., Wold, L. J., Gooding, L. R., and Wold, W. S. (1993) Characterization of mutants within the gene for the adenovirus E3 14.7-kilodalton protein which prevents cytolysis by tumor necrosis factor, *J. Virol. 67*, 2159–2167.

[40] Legrand, V., Spehner, D., Schlesinger, Y., Settelen, N., Pavirani, A., and Mehtali, M. (1999) Fiberless recombinant adenoviruses: virus maturation and infectivity in the absence of fiber, *J. Virol. 73*, 907–919.

[41] Von Seggern, D. J., Chiu, C. Y., Fleck, S. K., Stewart, P. L., and Nemerow, G. R. (1999) A helper-independent adenovirus vector with E1, E3, and fiber deleted: structure and infectivity of fiberless particles, *J. Virol. 73*, 1601–1608.

[42] Von Seggern, D. J., Kehler, J., Endo, R. I., and Nemerow, G. R. (1998) Complementation of a fibre mutant adenovirus by packaging cell lines stably expressing the adenovirus type 5 fibre protein, *J. Gen. Virol. 79* 1461–1468.

[43] Hodges, B. L., Serra, D., Hu, H., Begy, C. A., Chamberlain, J. S., and Amalfitano, A. (2000) Multiply deleted [E1, polymerase-, and pTP-] adenovirus vector persists despite deletion of the preterminal protein, *J. Gene Med. 2*, 250–259.

[44] Zhou, H., O'Neal, W., Morral, N., and Beaudet, A. L. (1996) Development of a complementing cell line and a system for construction of adenovirus vectors with E1 and E2a deleted, *J. Virol. 70*, 7030–7038.

[45] Rice, S. A., and Klessig, D. F. (1985) Isolation and analysis of adenovirus type 5 mutants containing deletions in the gene encoding the DNA-binding protein, *J. Virol. 56*, 767–778.

[46] Uil, T. G., Vellinga, J., de Vrij, J., van den Hengel, S. K., Rabelink, M. J., Cramer, S. J., Eekels, J. J., Ariyurek, Y., van Galen, M., and Hoeben, R. C. (2011) Directed adenovirus evolution using engineered mutator viral polymerases, *Nucleic Acids Res. 39*, e30.

[47] Hodges, B. L., Evans, H. K., Everett, R. S., Ding, E. Y., Serra, D., and Amalfitano, A. (2001) Adenovirus vectors with the 100K gene deleted and their potential for multiple gene therapy applications, *J. Virol. 75*, 5913–5920.

[48] Doronin, K., Toth, K., Kuppuswamy, M., Krajcsi, P., Tollefson, A. E., and Wold, W. S. (2003) Overexpression of the ADP (E3-11.6K) protein increases cell lysis and spread of adenovirus, *Virology 305*, 378–387.

[49] Tollefson, A. E., Ying, B., Doronin, K., Sidor, P. D., and Wold, W. S. (2007) Identification of a new human adenovirus protein encoded by a novel late l-strand transcription unit, *J. Virol. 81*, 12918–12926.

[50] Wang, H., Bian, X., Xia, L., Ding, X., Muller, R., Zhang, Y., Fu, J., and Stewart, A. F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection, *Nucleic Acids Res. 42*, e37.

[51] Papa, L. J., III, and Shoulders, M. D. (2019) Genetic Engineering by DNA Recombineering, *Curr. Protoc. Chem. Biol. 11*, e70.

[52] Kim, E., Ilic, N., Shrestha, Y., Zou, L., Kamburov, A., Zhu, C., Yang, X., Lubonja, R., Tran, N., Nguyen, C., Lawrence, M. S., Piccioni, F., Bagul, M., Doench, J. G., Chouinard, C. R., Wu, X., Hogstrom, L., Natoli, T., Tamayo, P., Horn, H., Corsello, S. M., Lage, K., Root, D. E., Subramanian, A., Golub, T. R., Getz, G., Boehm, J. S., and Hahn, W. C. (2016) Systematic functional interrogation of rare cancer variants identifies oncogenic alleles, *Cancer Discov. 6*, 714–726.

[53] Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H., and Moon, R. T. (2003) Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements, *Curr. Biol. 13*, 680–685.

[54] Näsvall, J. (2017) Direct and Inverted Repeat stimulated excision (DIRex): Simple, single-step, and scar-free mutagenesis of bacterial genes, *PLoS One 12*, e0184126.

Appendix B: Sequences

B.1 Chapter 2 sequences:

>pCDNA3.1-rtTA

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAG
TATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTT
GACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGC
GTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGAC
GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCAT
TATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT
GATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACG
CAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTA
CTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGTACCGA
GGATCTATTTCCGGTGAATTCCTCGAGGCCACCATGTCTAGACTGGACAAGAGCAAAGTCATAAACTCTGCTCTGGA
ATTACTCAATGGAGTCGGTATCGAAGGCCTGACGACAAGGAAACTCGCTCAAAAGCTGGGAGTTGAGCAGCCTACCC
TGTACTGGCACGTGAAGAACAAGCGGGCCCTGCTCGATGCCCTGCCAATCGAGATGCTGGACAGGCATCATACCCAC
TCCTGCCCCCTGGAAGGCGAGTCATGGCAAGACTTTCTGCGGAACAACGCCAAGTCATACCGCTGTGCTCTCCTCTC
ACATCGCGACGGGGCTAAAGTGCATCTCGGCACCCGCCCAACAGAGAAACAGTACGAAACCCTGGAAAATCAGCTCG
CGTTCCTGTGTCAGCAAGGCTTCTCCCTGGAGAACGCACTGTACGCTCTGTCCGCCGTGGGCCACTTTACACTGGGC
TGCGTATTGGAGGAACAGGAGCATCAAGTAGCAAAAGAGGAAAGAGAGACACCTACCACCGATTCTATGCCCCCACT
TCTGAAACAAGCAATTGAGCTGTTCGACCGGCAGGGAGCCGAACCTGCCTTCCTTTTCGGCCTGGAACTAATCATAT
GTGGCCTGGAGAAACAGCTAAAGTGCGAAAGCGGCGGGCCGACCGACGCCCTTGACGATTTTGACTTAGACATGCTC
CCAGCCGATGCCCTTGACGACTTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGACCTTGACATGCT
CCCCGGGTAAGGCGGCCGCATCGATAAGCTTGTCGACGATATCTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCT
CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACT
CCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTT
CTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT
GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTT
TCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC
GGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGC
CCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGT
CTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTA
ACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGC
AAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGC
ATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGC
CCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCC
AGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAACCTCCGCGGGGATCCGCACCATGGTGAGCAAGG
GCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCAC
GAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGG
TGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCG
CCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC
GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAA
CTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGG
ACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACC
ACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAA
CGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACA
AGTAGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAA
AGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAG
CATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACC
TCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAA
CATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCT
CACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGT
TTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATC

AGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCAC
AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGC
TTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCC
CCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCC
ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGC
CTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT
GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAA
AGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTG
CTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAG
CGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTC
GCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTT
CATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTC
GGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATC
ATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCG
TGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCG
CAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTAT
CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGACGTC

> pCDNA3.1-TTR

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAG
TATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTT
GACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGC
GTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGAC
GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCAT
TATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT
GATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACG
CAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTA
CTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTTGGTACCGAGCTCGGATCCGCCATGGCTTCT
CATCGTCTGCTCCTCCTCTGCCTTGCTGGACTGGTATTTGTGTCTGAGGCTGACTACAAAGACGATGACGATAAGGA
TTATAAGGATGACGATGACAAGGGTCCTACGGGCACCGGTGAATCCAAGTGTCCTCTGATGGTCAAAGTTCTAGATG
CTGTCCGAGGCAgTCCTGCCATCAATGTGgccgtgcatgtGTTCAGAAAGGCTGCTGATGACACCTGGGAGCCATTT
GCcTCTGGGAAAACCAGTGAGTCTGGAGAGCTGCATGGGCTCACAACTGAGGAGGAATTTGTAGAAGGGATATACAA
AGTGGAAATAGACACCAAATCTTACTGGAAGGCACTTGGCATCTCCCCATTCCATGAGCATGCAGAGGTGGTATTCA
CAGCCAACGACTCCGGCCCCCGCCGCTACACCATTGCCGCCCTGCTGAGCCCCTACTCCTATTCCACCACGGCTGTC
GTCACCAATCCCAAGGAATGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCT
ATTCTATAGTGTCACCTAAATGCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTG
CATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGG
GTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTG
CCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCT
CTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGA
TGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTG
GACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATT
TCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTA
GGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTG

TGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGC
CCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTT
ATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGG
CTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCA
TGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAA
CAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGA
CCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCG
CAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTG
TCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGC
TACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATC
AGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGAC
GGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATT
CATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGC
TTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTAT
CGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCA
CGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGAT
CCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAAT
AAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATC
AATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT
GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAAT
GAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCAT
TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
GCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGG
ATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTT
TTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACT
ATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTC
GTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATG
TAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTT
TTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT
TTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA
CGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTA
TCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTAT
TAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA
TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC
CCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATC
ACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGT
ACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC
GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACC
GCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTT
CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA
CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA
GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

> Ad.LacZ

CATCATCAATAATATACCTTATTTTGGATTGAAGCCAATATGATAATGAGGGGGTGGAGTTTGTGACGTGGCGCGGG
GCGTGGGAACGGGGCGGGTGACGTAGTAGTGTGGCGGAAGTGTGATGTTGCAAGTGTGGCGGAACACATGTAAGCGA
CGGATGTGGCAAAAGTGACGTTTTTGGTGTGCGCCGGTGTACACAGGAAGTGACAATTTTCGCGCGGTTTTAGGCGG
ATGTTGTAGTAAATTTGGGCGTAACCGAGTAAGATTTGGCCATTTTCGCGGGAAAACTGAATAAGAGGAAGTGAAAT
CTGAATAATTTTGTGTTACTCATAGCGCGTAATAGAAGCCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGT
CTTCCCAATCCTCCCCCTTGCTGTCCTGCCCCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGAT
GCAATTTCCTCATTTTATTAGGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCA

AACAACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGTTTAAACGGGCCCTCTAGAGATATCGT
CGACAAGCTTATCGATGCGGCCGCTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAA
TTTCACAGGAGGTACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACC
CTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACC
GATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATAGGGATCCAATTCTTGAAGACGAAAGGGCCTCGTGAT
ACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGC
CCGGTTAACGTGCCGGCACGGCCTGGGTAACCAGGTATTTTGTCCACATAACCGTGCGCAAAATGTTGTGGATAAGC
AGGACACAGCAGCAATCCACAGCAGGCATACAACCGCACACCGAGGTTACTCCGTTCTACAGGTTACGACGACATGT
CAATACTTGCCCTTGACAGGCATTGATGGAATCGTAGTCTCACGCTGAGGTACCAAGCTTAAGTTTAAACGCTAGCC
AGCTTGGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGCCAGTAAGCAGTGGGTTCTCTAGTTAGCCAGAGAG
CTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTT
GGAAAGTCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGA
GTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGACTAATACGTAGA
TGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACG
TCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCA
TTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGT
TGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCGGAACTCCATATATGGGCTATGAACTAATGAC
CCCGTAATTGATTACTATTAATAACTAGTCAATAATCAATGTCAACGCGTATATCTGTGGGCGTGGCTTAAGGGTGG
GAAAGAATATATAAGGTGGGGGTCTTATGTAGTTTTGTATCTGTTTTGCAGCAGCCGCCGCCGCCATGAGCACCAAC
TCGTTTGATGGAAGCATTGTGAGCTCATATTTGACAACGCGCATGCCCCCATGGGCCGGGGTGCGTCAGAATGTGAT
GGGCTCCAGCATTGATGGTCGCCCCGTCCTGCCCGCAAACTCTACTACCTTGACCTACGAGACCGTGTCTGGAACGC
CGTTGGAGACTGCAGCCTCCGCCGCCGCTTCAGCCGCTGCAGCCACCGCCCGCGGGATTGTGACTGACTTTGCTTTC
CTGAGCCCGCTTGCAAGCAGTGCAGCTTCCCGTTCATCCGCCCGCGATGACAAGTTGACGGCTCTTTTGGCACAATT
GGATTCTTTGACCCGGGAACTTAATGTCGTTTCTCAGCAGCTGTTGGATCTGCGCCAGCAGGTTTCTGCCCTGAAGG
CTTCCTCCCCTCCCAATGCGGTTTAAAACATAAATAAAAAACCAGACTCTGTTTGGATTTGGATCAAGCAAGTGTCT
TGCTGTCTTTATTTAGGGGTTTTGCGCGCGCGGTAGGCCCGGGACCAGCGGTCTCGGTCGTTGAGGGTCCTGTGTAT
TTTTTCCAGGACGTGGTAAAGGTGACTCTGGATGTTCAGATACATGGGCATAAGCCCGTCTCTGGGGTGGAGGTAGC
ACCACTGCAGAGCTTCATGCTGCGGGGTGGTGTTGTAGATGATCCAGTCGTAGCAGGAGCGCTGGGCGTGGTGCCTA
AAAATGTCTTTCAGTAGCAAGCTGATTGCCAGGGGCAGGCCCTTGGTGTAAGTGTTTACAAAGCGGTTAAGCTGGGA
TGGGTGCATACGTGGGGATATGAGATGCATCTTGGACTGTATTTTTAGGTTGGCTATGTTCCCAGCCATATCCCTCC
GGGGATTCATGTTGTGCAGAACCACCAGCACAGTGTATCCGGTGCACTTGGGAAATTTGTCATGTAGCTTAGAAGGA
AATGCGTGGAAGAACTTGGAGACGCCCTTGTGACCTCCAAGATTTTCCATGCATTCGTCCATAATGATGGCAATGGG
CCCACGGGCGGCGGCCTGGGCGAAGATATTTCTGGGATCACTAACGTCATAGTTGTGTTCCAGGATGAGATCGTCAT
AGGCCATTTTTACAAAGCGCGGGCGGAGGGTGCCAGACTGCGGTATAATGGTTCCATCCGGCCCAGGGGCGTAGTTA
CCCTCACAGATTTGCATTTCCCACGCTTTGAGTTCAGATGGGGGGATCATGTCTACCTGCGGGGCGATGAAGAAAAC
GGTTTCCGGGGTAGGGGAGATCAGCTGGGAAGAAAGCAGGTTCCTGAGCAGCTGCGACTTACCGCAGCCGGTGGGCC
CGTAAATCACACCTATTACCGGGTGCAACTGGTAGTTAAGAGAGCTGCAGCTGCCGTCATCCCTGAGCAGGGGGGCC
ACTTCGTTAAGCATGTCCCTGACTCGCATGTTTTCCCTGACCAAATCCGCCAGAAGGCGCTCGCCGCCCAGCGATAG
CAGTTCTTGCAAGGAAGCAAAGTTTTTCAACGGTTTGAGACCGTCCGCCGTAGGCATGCTTTTGAGCGTTTGACCAA
GCAGTTCCAGGCGGTCCCACAGCTCGGTCACCTGCTCTACGGCATCTCGATCCAGCATATCTCCTCGTTTCGCGGGT
TGGGGCGGCTTTCGCTGTACGGCAGTAGTCGGTGCTCGTCCAGACGGGCCAGGGTCATGTCTTTCCACGGGCGCAGG
GTCCTCGTCAGCGTAGTCTGGGTCACGGTGAAGGGGTGCGCTCCGGGCTGCGCGCTGGCCAGGGTGCGCTTGAGGCT
GGTCCTGCTGGTGCTGAAGCGCTGCCGGTCTTCGCCCTGCGCGTCGGCCAGGTAGCATTTGACCATGGTGTCATAGT
CCAGCCCCTCCGCGGCGTGGCCCTTGGCGCGCAGCTTGCCCTTGGAGGAGGCGCCGCACGAGGGGCAGTGCAGACTT
TTGAGGGCGTAGAGCTTGGGCGCGAGAAATACCGATTCCGGGGAGTAGGCATCCGCGCCGCAGGCCCCGCAGACGGT
CTCGCATTCCACGAGCCAGGTGAGCTCTGGCCGTTCGGGGTCAAAAACCAGGTTTCCCCCATGCTTTTTGATGCGTT
TCTTACCTCTGGTTTCCATGAGCCGGTGTCCACGCTCGGTGACGAAAAGGCTGTCCGTGTCCCCGTATACAGACTTG
AGAGGCCTGTCCTCGAGCGGTGTTCCGCGGTCCTCCTCGTATAGAAACTCGGACCACTCTGAGACAAAGGCTCGCGT
CCAGGCCAGCACGAAGGAGGCTAAGTGGGAGGGGTAGCGGTCGTTGTCCACTAGGGGGTCCACTCGCTCCAGGGTGT
GAAGACACATGTCGCCCTCTTCGGCATCAAGGAAGGTGATTGGTTTGTAGGTGTAGGCCACGTGACCGGGTGTTCCT
GAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCACTCTCTTCCGCATCGCTGTCTGCGAGGGCCAGCTG
TTGGGGTGAGTACTCCCTCTGAAAAGCGGGCATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAAACGAGGAGGATT
TGATATTCACCTGGCCCGCGGTGATGCCTTTGAGGGTGGCCGCATCCATCTGGTCAGAAAAGACAATCTTTTTGTTG
TCAAGCTTGGTGGCAAACGACCCGTAGAGGGCGTTGGACAGCAACTTGGCGATGGAGCGCAGGGTTTGGTTTTTGTC
GCGATCGGCGCGCTCCTTGGCCGCGATGTTTAGCTGCACGTATTCGCGCGCAACGCACCGCCATTCGGGAAAGACGG
TGGTGCGCTCGTCGGGCACCAGGTGCACGCGCCAACCGCGGTTGTGCAGGGTGACAAGGTCAACGCTGGTGGCTACC
TCTCCGCGTAGGCGCTCGTTGGTCCAGCAGAGGCGGCCGCCCTTGCGCGAGCAGAATGGCGGTAGGGGGTCTAGCTG

CGTCTCGTCCGGGGGGTCTGCGTCCACGGTAAAGACCCCGGGCAGCAGGCGCGCGTCGAAGTAGTCTATCTTGCATC
CTTGCAAGTCTAGCGCCTGCTGCCATGCGCGGGCGGCAAGCGCGCGCTCGTATGGGTTGAGTGGGGGACCCCATGGC
ATGGGGTGGGTGAGCGCGGAGGCGTACATGCCGCAAATGTCGTAAACGTAGAGGGGCTCTCTGAGTATTCCAAGATA
TGTAGGGTAGCATCTTCCACCGCGGATGCTGGCGCGCACGTAATCGTATAGTTCGTGCGAGGGAGCGAGGAGGTCGG
GACCGAGGTTGCTACGGGCGGGCTGCTCTGCTCGGAAGACTATCTGCCTGAAGATGGCATGTGAGTTGGATGATATG
GTTGGACGCTGGAAGACGTTGAAGCTGGCGTCTGTGAGACCTACCGCGTCACGCACGAAGGAGGCGTAGGAGTCGCG
CAGCTTGTTGACCAGCTCGGCGGTGACCTGCACGTCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACT
TATCCTGTCCCTTTTTTTTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGTACTCTTGGATCGGA
AACCCGTCGGCCTCCGAACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTGGTAGGCGCAGCATCCCTTTTC
TACGGGTAGCGCGTATGCCTGCGCGGCCTTCCGGAGCGAGGTGTGGGTGAGCGCAAAGGTGTCCCTGACCATGACCA
GCATGAAGGGCACGAGCTGCTTCCCAAAGGCCCCCATCCAAGTATAGGTCTCTACATCGTAGGTGACAAAGAGACGC
TCGGTGCGAGGATGCGAGCCGATCGGGAAGAACTGGATCTCCCGCCACCAATTGGAGGAGTGGCTATTGATGTGGTG
AAAGTAGAAGTCCCTGCGACGGGCCGAACACTCGTGCTGGCTTTTGTAAAAACGTGCGCAGTACTGGCAGCGGTGCA
CGGGCTGTACATCCTGCACGAGGTTGACCTGACGACCGCGCACAAGGAAGCAGAGTGGGAATTTGAGCCCCTCGCCT
GGCGGGTTTGGCTGGTGGTCTTCTACTTCGGCTGCTTGTCCTTGACCGTCTGGCTGCTCGAGGGGAGTTACGGTGGA
TCGGACCACCACGCCGCGCGAGCCCAAAGTCCAGATGTCCGCGCGCGGCGGTCGGAGCTTGATGACAACATCGCGCA
GATGGGAGCTGTCCATGGTCTGGAGCTCCCGCGGCGTCAGGTCAGGCGGGAGCTCCTGCAGGTTTACCTCGCATAGA
CGGGTCAGGGCGCGGGCTAGATCCAGGTGATACCTAATTTCCAGGGGCTGGTTGGTGGCGGCGTCGATGGCTTGCAA
GAGGCCGCATCCCCGCGGCGCGACTACGGTACCGCGCGGCGGGCGGTGGGCCGCGGGGGTGTCCTTGGATGATGCAT
CTAAAAGCGGTGACGCGGGCGAGCCCCCGGAGGTAGGGGGGGCTCCGGACCCGCCGGGAGAGGGGGCAGGGGCACGT
CGGCGCCGCGCGCGGGCAGGAGCTGGTGCTGCGCGCGTAGGTTGCTGGCGAACGCGACGACGCGGCGGTTGATCTCC
TGAATCTGGCGCCTCTGCGTGAAGACGACGGGCCCGGTGAGCTTGAGCCTGAAAGAGAGTTCGACAGAATCAATTTC
GGTGTCGTTGACGGCGGCCTGGCGCAAAATCTCCTGCACGTCTCCTGAGTTGTCTTGATAGGCGATCTCGGCCATGA
ACTGCTCGATCTCTTCCTCCTGGAGATCTCCGCGTCCGGCTCGCTCCACGGTGGCGGCGAGGTCGTTGGAAATGCGG
GCCATGAGCTGCGAGAAGGCGTTGAGGCCTCCCTCGTTCCAGACGCGGCTGTAGACCACGCCCCCTTCGGCATCGCG
GGCGCGCATGACCACCTGCGCGAGATTGAGCTCCACGTGCCGGGCGAAGACGGCGTAGTTTCGCAGGCGCTGAAAGA
GGTAGTTGAGGGTGGTGGCGGTGTGTTCTGCCACGAAGAAGTACATAACCCAGCGTCGCAACGTGGATTCGTTGATA
TCCCCCAAGGCCTCAAGGCGCTCCATGGCCTCGTAGAAGTCCACGGCGAAGTTGAAAAACTGGGAGTTGCGCGCCGA
CACGGTTAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGACAGTGTCGCGCACCTCGCGCTCAAAGGCTACAGGGG
CCTCTTCTTCTTCTTCAATCTCCTCTTCCATAAGGGCCTCCCCTTCTTCTTCTTCTGGCGGCGGTGGGGGAGGGGGG
ACACGGCGGCGACGACGGCGCACCGGGAGGCGGTCGACAAAGCGCTCGATCATCTCCCCGCGGCGACGGCGCATGGT
CTCGGTGACGGCGCGGCCGTTCTCGCGGGGGCGCAGTTGGAAGACGCCGCCCGTCATGTCCCGGTTATGGGTTGGCG
GGGGGCTGCCATGCGGCAGGGATACGGCGCTAACGATGCATCTCAACAATTGTTGTGTAGGTACTCCGCCGCCGAGG
GACCTGAGCGAGTCCGCATCGACCGGATCGGAAAACCTCTCGAGAAAGGCGTCTAACCAGTCACAGTCGCAAGGTAG
GCTGAGCACCGTGGCGGGCGGCAGCGGGCGGCGGTCGGGGTTGTTTCTGGCGGAGGTGCTGCTGATGATGTAATTAA
AGTAGGCGGTCTTGAGACGGCGGATGGTCGACAGAAGCACCATGTCCTTGGGTCCGGCCTGCTGAATGCGCAGGCGG
TCGGCCATGCCCCAGGCTTCGTTTTGACATCGGCGCAGGTCTTTGTAGTAGTCTTGCATGAGCCTTTCTACCGGCAC
TTCTTCTTCTCCTTCCTCTTGTCCTGCATCTCTTGCATCTATCGCTGCGGCGGCGGCGGAGTTTGGCCGTAGGTGGC
GCCCTCTTCCTCCCATGCGTGTGACCCCGAAGCCCCTCATCGGCTGAAGCAGGGCTAGGTCGGCGACAACGCGCTCG
GCTAATATGGCCTGCTGCACCTGCGTGAGGGTAGACTGGAAGTCATCCATGTCCACAAAGCGGTGGTATGCGCCCGT
GTTGATGGTGTAAGTGCAGTTGGCCATAACGGACCAGTTAACGGTCTGGTGACCCGGCTGCGAGAGCTCGGTGTACC
TGAGACGCGAGTAAGCCCTCGAGTCAAATACGTAGTCGTTGCAAGTCCGCACCAGGTACTGGTATCCCACCAAAAAG
TGCGGCGGCGGCTGGCGGTAGAGGGGCCAGCGTAGGGTGGCCGGGGCTCCGGGGGCGAGATCTTCCAACATAAGGCG
ATGATATCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGAAAGTCGCGGACGC
GGTTCCAGATGTTGCGCAGCGGCAAAAAGTGCTCCATGGTCGGGACGCTCTGGCCGGTCAGGCGCGCGCAATCGTTG
ACGCTCTAGACCGTGCAAAAGGAGAGCCTGTAAGCGGGCACTCTTCCGTGGTCTGGTGGATAAATTCGCAAGGGTAT
CATGGCGGACGACCGGGGTTCGAGCCCCGTATCCGGCCGTCCGCCGTGATCCATGCGGTTACCGCCCGCGTGTCGAA
CCCAGGTGTGCGACGTCAGACAACGGGGGAGTGCTCCTTTTGGCTTCCTTCCAGGCGCGGCGGCTGCTGCGCTAGCT
TTTTTGGCCACTGGCCGCGCGCAGCGTAAGCGGTTAGGCTGGAAAGCGAAAGCATTAAGTGGCTCGCTCCCTGTAGC
CGGAGGGTTATTTTCCAAGGGTTGAGTCGCGGGACCCCCGGTTCGAGTCTCGGACCGGCCGGACTGCGGCGAACGGG
GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAACAGGGACGAGCCCCTTTTTTGCTTTTCC
CAGATGCATCCGGTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAGCAAGAGCAGCGGCAGACATGCAG
GGCACCCTCCCCTCCTCCTACCGCGTCAGGAGGGGCGACATCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAAC
CCCCGCGGCGCCGGGCCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCTAGGAGCGCCCTCTCCT
GAGCGGTACCCAAGGGTGCAGCTGAAGCGTGATACGCGTGAGGCGTACGTGCCGCGGCAGAACCTGTTTCGCGACCG
CGAGGGAGAGGAGCCCGAGGAGATGCGGGATCGAAAGTTCCACGCAGGGCGCGAGCTGCGGCATGGCCTGAATCGCG
AGCGGTTGCTGCGCGAGGAGGACTTTGAGCCCGACGCGCGAACCGGGATTAGTCCCGCGCGCGCACACGTGGCGGCC

GCCGACCTGGTAACCGCATACGAGCAGACGGTGAACCAGGAGATTAACTTTCAAAAAAGCTTTAACAACCACGTGCG
TACGCTTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTTGTAAGCGCGCTGGAGCAAAACC
CAAATAGCAAGCCGCTCATGGCGCAGCTGTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCG
CTGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTCGATTTGATAAACATCCTGCAGAGCATAGTGGTGCAGGA
GCGCAGCTTGAGCCTGGCTGACAAGGTGGCCGCCATCAACTATTCCATGCTTAGCCTGGGCAAGTTTTACGCCCGCA
AGATATACCATACCCCTTACGTTCCCATAGACAAGGAGGTAAAGATCGAGGGGTTCTACATGCGCATGGCGCTGAAG
GTGCTTACCTTGAGCGACGACCTGGGCGTTTATCGCAACGAGCGCATCCACAAGGCCGTGAGCGTGAGCCGGCGGCG
CGAGCTCAGCGACCGCGAGCTGATGCACAGCCTGCAAAGGGCCCTGGCTGGCACGGGCAGCGGCGATAGAGAGGCCG
AGTCCTACTTTGACGCGGGCGCTGACCTGCGCTGGGCCCCAAGCCGACGCGCCCTGGAGGCAGCTGGGGCCGGACCT
GGGCTGGCGGTGGCACCCGCGCGCGCTGGCAACGTCGGCGGCGTGGAGGAATATGACGAGGACGATGAGTACGAGCC
AGAGGACGGCGAGTACTAAGCGGTGATGTTTCTGATCAGATGATGCAAGACGCAACGGACCCGGCGGTGCGGGCGGC
GCTGCAGAGCCAGCCGTCCGGCCTTAACTCCACGGACGACTGGCGCCAGGTCATGGACCGCATCATGTCGCTGACTG
CGCGCAATCCTGACGCGTTCCGGCAGCAGCCGCAGGCCAACCGGCTCTCCGCAATTCTGGAAGCGGTGGTCCCGGCG
CGCGCAAACCCCACGCACGAGAAGGTGCTGGCGATCGTAAACGCGCTGGCCGAAAACAGGGCCATCCGGCCCGACGA
GGCCGGCCTGGTCTACGACGCGCTGCTTCAGCGCGTGGCTCGTTACAACAGCGGCAACGTGCAGACCAACCTGGACC
GGCTGGTGGGGGATGTGCGCGAGGCCGTGGCGCAGCGTGAGCGCGCGCAGCAGCAGGGCAACCTGGGCTCCATGGTT
GCACTAAACGCCTTCCTGAGTACACAGCCCGCCAACGTGCCGCGGGGACAGGAGGACTACACCAACTTTGTGAGCGC
ACTGCGGCTAATGGTGACTGAGACACCGCAAAGTGAGGTGTACCAGTCTGGGCCAGACTATTTTTTCCAGACCAGTA
GACAAGGCCTGCAGACCGTAAACCTGAGCCAGGCTTTCAAAAACTTGCAGGGGCTGTGGGGGGTGCGGGCTCCCACA
GGCGACCGCGCGACCGTGTCTAGCTTGCTGACGCCCAACTCGCGCCTGTTGCTGCTGCTAATAGCGCCCTTCACGGA
CAGTGGCAGCGTGTCCCGGGACACATACCTAGGTCACTTGCTGACACTGTACCGCGAGGCCATAGGTCAGGCGCATG
TGGACGAGCATACTTTCCAGGAGATTACAAGTGTCAGCCGCGCGCTGGGGCAGGAGGACACGGGCAGCCTGGAGGCA
ACCCTAAACTACCTGCTGACCAACCGGCGGCAGAAGATCCCCTCGTTGCACAGTTTAAACAGCGAGGAGGAGCGCAT
TTTGCGCTACGTGCAGCAGAGCGTGAGCCTTAACCTGATGCGCGACGGGGTAACGCCCAGCGTGGCGCTGGACATGA
CCGCGCGCAACATGGAACCGGGCATGTATGCCTCAAACCGGCCGTTTATCAACCGCCTAATGGACTACTTGCATCGC
GCGGCCGCCGTGAACCCCGAGTATTTCACCAATGCCATCTTGAACCCGCACTGGCTACCGCCCCCTGGTTTCTACAC
CGGGGGATTCGAGGTGCCCGAGGGTAACGATGGATTCCTCTGGGACGACATAGACGACAGCGTGTTTTCCCCGCAAC
CGCAGACCCTGCTAGAGTTGCAACAGCGCGAGCAGGCAGAGGCGGCGCTGCGAAAGGAAAGCTTCCGCAGGCCAAGC
AGCTTGTCCGATCTAGGCGCTGCGGCCCCGCGGTCAGATGCTAGTAGCCCATTTCCAAGCTTGATAGGGTCTCTTAC
CAGCACTCGCACCACCCGCCCGCGCCTGCTGGGCGAGGAGGAGTACCTAAACAACTCGCTGCTGCAGCCGCAGCGCG
AAAAAAACCTGCCTCCGGCATTTCCCAACAACGGGATAGAGAGCCTAGTGGACAAGATGAGTAGATGGAAGACGTAC
GCGCAGGAGCACAGGGACGTGCCAGGCCCGCGCCCGCCCACCCGTCGTCAAAGGCACGACCGTCAGCGGGGTCTGGT
GTGGGAGGACGATGACTCGGCAGACGACAGCAGCGTCCTGGATTTGGGAGGGAGTGGCAACCCGTTTGCGCACCTTC
GCCCCAGGCTGGGGAGAATGTTTTAAAAAAAAAAAAGCATGATGCAAAATAAAAAACTCACCAAGGCCATGGCACCG
AGCGTTGGTTTTCTTGTATTCCCCTTAGTATGCGGCGCGCGGCGATGTATGAGGAAGGTCCTCCTCCCTCCTACGAG
AGTGTGGTGAGCGCGGCGCCAGTGGCGGCGGCGCTGGGTTCTCCCTTCGATGCTCCCCTGGACCCGCCGTTTGTGCC
TCCGCGGTACCTGCGGCCTACCGGGGGGAGAAACAGCATCCGTTACTCTGAGTTGGCACCCCTATTCGACACCACCC
GTGTGTACCTGGTGGACAACAAGTCAACGGATGTGGCATCCCTGAACTACCAGAACGACCACAGCAACTTTCTGACC
ACGGTCATTCAAAACAATGACTACAGCCCGGGGGAGGCAAGCACACAGACCATCAATCTTGACGACCGGTCGCACTG
GGGCGGCGACCTGAAAACCATCCTGCATACCAACATGCCAAATGTGAACGAGTTCATGTTTACCAATAAGTTTAAGG
CGCGGGTGATGGTGTCGCGCTTGCCTACTAAGGACAATCAGGTGGAGCTGAAATACGAGTGGGTGGAGTTCACGCTG
CCCGAGGGCAACTACTCCGAGACCATGACCATAGACCTTATGAACAACGCGATCGTGGAGCACTACTTGAAAGTGGG
CAGACAGAACGGGGTTCTGGAAAGCGACATCGGGGTAAAGTTTGACACCCGCAACTTCAGACTGGGGTTTGACCCCG
TCACTGGTCTTGTCATGCCTGGGGTATATACAAACGAAGCCTTCCATCCAGACATCATTTTGCTGCCAGGATGCGGG
GTGGACTTCACCCACAGCCGCCTGAGCAACTTGTTGGGCATCCGCAAGCGGCAACCCTTCCAGGAGGGCTTTAGGAT
CACCTACGATGATCTGGAGGGTGGTAACATTCCCGCACTGTTGGATGTGGACGCCTACCAGGCGAGCTTGAAAGATG
ACACCGAACAGGGCGGGGGTGGCGCAGGCGGCAGCAACAGCAGTGGCAGCGGCGCGGAAGAGAACTCCAACGCGGCA
GCCGCGGCAATGCAGCCGGTGGAGGACATGAACGATCATGCCATTCGCGGCGACACCTTTGCCACACGGGCTGAGGA
GAAGCGCGCTGAGGCCGAAGCAGCGGCCGAAGCTGCCGCCCCCGCTGCGCAACCCGAGGTCGAGAAGCCTCAGAAGA
AACCGGTGATCAAACCCCTGACAGAGGACAGCAAGAAACGCAGTTACAACCTAATAAGCAATGACAGCACCTTCACC
CAGTACCGCAGCTGGTACCTTGCATACAACTACGGCGACCCTCAGACCGGAATCCGCTCATGGACCCTGCTTTGCAC
TCCTGACGTAACCTGCGGCTCGGAGCAGGTCTACTGGTCGTTGCCAGACATGATGCAAGACCCCGTGACCTTCCGCT
CCACGCGCCAGATCAGCAACTTTCCGGTGGTGGGCGCCGAGCTGTTGCCCGTGCACTCCAAGAGCTTCTACAACGAC
CAGGCCGTCTACTCCCAACTCATCCGCCAGTTTACCTCTCTGACCCACGTGTTCAATCGCTTTCCCGAGAACCAGAT
TTTGGCGCGCCCGCCAGCCCCCACCATCACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACGCTAC
CGCTGCGCAACAGCATCGGAGGAGTCCAGCGAGTGACCATTACTGACGCCAGACGCCGCACCTGCCCCTACGTTTAC
AAGGCCCTGGGCATAGTCTCGCCGCGCGTCCTATCGAGCCGCACTTTTTGAGCAAGCATGTCCATCCTTATATCGCC

CAGCAATAACACAGGCTGGGGCCTGCGCTTCCCAAGCAAGATGTTTGGCGGGGCCAAGAAGCGCTCCGACCAACACC
CAGTGCGCGTGCGCGGGCACTACCGCGCGCCCTGGGGCGCGCACAAACGCGGCCGCACTGGGCGCACCACCGTCGAT
GACGCCATCGACGCGGTGGTGGAGGAGGCGCGCAACTACACGCCCACGCCGCCACCAGTGTCCACAGTGGACGCGGC
CATTCAGACCGTGGTGCGCGGAGCCCGGCGCTATGCTAAAATGAAGAGACGGCGGAGGCGCGTAGCACGTCGCCACC
GCCGCCGACCCGGCACTGCCGCCCAACGCGCGGCGGCGGCCCTGCTTAACCGCGCACGTCGCACCGGCCGACGGGCG
GCCATGCGGGCCGCTCGAAGGCTGGCCGCGGGTATTGTCACTGTGCCCCCCAGGTCCAGGCGACGAGCGGCCGCCGC
AGCAGCCGCGGCCATTAGTGCTATGACTCAGGGTCGCAGGGGCAACGTGTATTGGGTGCGCGACTCGGTTAGCGGCC
TGCGCGTGCCCGTGCGCACCCGCCCCCCGCGCAACTAGATTGCAAGAAAAAACTACTTAGACTCGTACTGTTGTATG
TATCCAGCGGCGGCGGCGCGCAACGAAGCTATGTCCAAGCGCAAAATCAAAGAAGAGATGCTCCAGGTCATCGCGCC
GGAGATCTATGGCCCCCCGAAGAAGGAAGAGCAGGATTACAAGCCCCGAAAGCTAAAGCGGGTCAAAAAGAAAAAGA
AAGATGATGATGATGAACTTGACGACGAGGTGGAACTGCTGCACGCTACCGCGCCCAGGCGACGGGTACAGTGGAAA
GGTCGACGCGTAAAACGTGTTTTGCGACCCGGCACCACCGTAGTCTTTACGCCCGGTGAGCGCTCCACCCGCACCTA
CAAGCGCGTGTATGATGAGGTGTACGGCGACGAGGACCTGCTTGAGCAGGCCAACGAGCGCCTCGGGGAGTTTGCCT
ACGGAAAGCGGCATAAGGACATGCTGGCGTTGCCGCTGGACGAGGGCAACCCAACACCTAGCCTAAAGCCCGTAACA
CTGCAGCAGGTGCTGCCCGCGCTTGCACCGTCCGAAGAAAAGCGCGGCCTAAAGCGCGAGTCTGGTGACTTGGCACC
CACCGTGCAGCTGATGGTACCCAAGCGCCAGCGACTGGAAGATGTCTTGGAAAAAATGACCGTGGAACCTGGGCTGG
AGCCCGAGGTCCGCGTGCGGCCAATCAAGCAGGTGGCGCCGGGACTGGGCGTGCAGACCGTGGACGTTCAGATACCC
ACTACCAGTAGCACCAGTATTGCCACCGCCACAGAGGGCATGGAGACACAAACGTCCCCGGTTGCCTCAGCGGTGGC
GGATGCCGCGGTGCAGGCGGTCGCTGCGGCCGCGTCCAAGACCTCTACGGAGGTGCAAACGGACCCGTGGATGTTTC
GCGTTTCAGCCCCCCGGCGCCCGCGCGGTTCGAGGAAGTACGGCGCCGCCAGCGCGCTACTGCCCGAATATGCCCTA
CATCCTTCCATTGCGCCTACCCCCGGCTATCGTGGCTACACCTACCGCCCCAGAAGACGAGCAACTACCCGACGCCG
AACCACCACTGGAACCCGCCGCCGCCGTCGCCGTCGCCAGCCCGTGCTGGCCCCGATTTCCGTGCGCAGGGTGGCTC
GCGAAGGAGGCAGGACCCTGGTGCTGCCAACAGCGCGCTACCACCCCAGCATCGTTTAAAAGCCGGTCTTTGTGGTT
CTTGCAGATATGGCCCTCACCTGCCGCCTCCGTTTCCCGGTGCCGGGATTCCGAGGAAGAATGCACCGTAGGAGGGG
CATGGCCGGCCACGGCCTGACGGGCGGCATGCGTCGTGCGCACCACCGGCGGCGGCGCGCGTCGCACCGTCGCATGC
GCGGCGGTATCCTGCCCCTCCTTATTCCACTGATCGCCGCGGCGATTGGCGCCGTGCCCGGAATTGCATCCGTGGCC
TTGCAGGCGCAGAGACACTGATTAAAAACAAGTTGCATGTGGAAAAATCAAAATAAAAAGTCTGGACTCTCACGCTC
GCTTGGTCCTGTAACTATTTTGTAGAATGGAAGACATCAACTTTGCGTCTCTGGCCCCGCGACACGGCTCGCGCCCG
TTCATGGGAAACTGGCAAGATATCGGCACCAGCAATATGAGCGGTGGCGCCTTCAGCTGGGGCTCGCTGTGGAGCGG
CATTAAAAATTTCGGTTCCACCGTTAAGAACTATGGCAGCAAGGCCTGGAACAGCAGCACAGGCCAGATGCTGAGGG
ATAAGTTGAAAGAGCAAAATTTCCAACAAAAGGTGGTAGATGGCCTGGCCTCTGGCATTAGCGGGGTGGTGGACCTG
GCCAACCAGGCAGTGCAAAATAAGATTAACAGTAAGCTTGATCCCCGCCCTCCCGTAGAGGAGCCTCCACCGGCCGT
GGAGACAGTGTCTCCAGAGGGGCGTGGCGAAAAGCGTCCGCGCCCCGACAGGGAAGAAACTCTGGTGACGCAAATAG
ACGAGCCTCCCTCGTACGAGGAGGCACTAAAGCAAGGCCTGCCCACCACCCGTCCCATCGCGCCCATGGCTACCGGA
GTGCTGGGCCAGCACACACCCGTAACGCTGGACCTGCCTCCCCCCGCCGACACCCAGCAGAAACCTGTGCTGCCAGG
CCCGACCGCCGTTGTTGTAACCCGTCCTAGCCGCGCGTCCCTGCGCCGCGCCGCCAGCGGTCCGCGATCGTTGCGGC
CCGTAGCCAGTGGCAACTGGCAAAGCACACTGAACAGCATCGTGGGTCTGGGGGTGCAATCCCTGAAGCGCCGACGA
TGCTTCTGAATAGCTAACGTGTCGTATGTGTGTCATGTATGCGTCCATGTCGCCGCCAGAGGAGCTGCTGAGCCGCC
GCGCGCCCGCTTTCCAAGATGGCTACCCCTTCGATGATGCCGCAGTGGTCTTACATGCACATCTCGGGCCAGGACGC
CTCGGAGTACCTGAGCCCCGGGCTGGTGCAGTTTGCCCGCGCCACCGAGACGTACTTCAGCCTGAATAACAAGTTTA
GAAACCCCACGGTGGCGCCTACGCACGACGTGACCACAGACCGGTCCCAGCGTTTGACGCTGCGGTTCATCCCTGTG
GACCGTGAGGATACTGCGTACTCGTACAAGGCGCGGTTCACCCTAGCTGTGGGTGATAACCGTGTGCTGGACATGGC
TTCCACGTACTTTGACATCCGCGGCGTGCTGGACAGGGGCCCTACTTTTAAGCCCTACTCTGGCACTGCCTACAACG
CCCTGGCTCCCAAGGGTGCCCCAAATCCTTGCGAATGGGATGAAGCTGCTACTGCTCTTGAAATAAACCTAGAAGAA
GAGGACGATGACAACGAAGACGAAGTAGACGAGCAAGCTGAGCAGCAAAAAACTCACGTATTTGGGCAGGCGCCTTA
TTCTGGTATAAATATTACAAAGGAGGGTATTCAAATAGGTGTCGAAGGTCAAACACCTAAATATGCCGATAAAACAT
TTCAACCTGAACCTCAAATAGGAGAATCTCAGTGGTACGAAACTGAAATTAATCATGCAGCTGGGAGAGTCCTTAAA
AAGACTACCCCAATGAAACCATGTTACGGTTCATATGCAAAACCCACAAATGAAAATGGAGGGCAAGGCATTCTTGT
AAAGCAACAAAATGGAAAGCTAGAAAGTCAAGTGGAAATGCAATTTTTCTCAACTACTGAGGCGACCGCAGGCAATG
GTGATAACTTGACTCCTAAAGTGGTATTGTACAGTGAAGATGTAGATATAGAAACCCCAGACACTCATATTTCTTAC
ATGCCCACTATTAAGGAAGGTAACTCACGAGAACTAATGGGCCAACAATCTATGCCCAACAGGCCTAATTACATTGC
TTTTAGGGACAATTTTATTGGTCTAATGTATTACAACAGCACGGGTAATATGGGTGTTCTGGCGGGCCAAGCATCGC
AGTTGAATGCTGTTGTAGATTTGCAAGACAGAAACACAGAGCTTTCATACCAGCTTTTGCTTGATTCCATTGGTGAT
AGAACCAGGTACTTTTCTATGTGGAATCAGGCTGTTGACAGCTATGATCCAGATGTTAGAATTATTGAAAATCATGG
AACTGAAGATGAACTTCCAAATTACTGCTTTCCACTGGGAGGTGTGATTAATACAGAGACTCTTACCAAGGTAAAAC
CTAAAACAGGTCAGGAAAATGGATGGGAAAAAGATGCTACAGAATTTTCAGATAAAAATGAAATAAGAGTTGGAAAT
AATTTTGCCATGGAAATCAATCTAAATGCCAACCTGTGGAGAAATTTCCTGTACTCCAACATAGCGCTGTATTTGCC

CGACAAGCTAAAGTACAGTCCTTCCAACGTAAAAATTTCTGATAACCCAAACACCTACGACTACATGAACAAGCGAG
TGGTGGCTCCCGGGTTAGTGGACTGCTACATTAACCTTGGAGCACGCTGGTCCCTTGACTATATGGACAACGTCAAC
CCATTTAACCACCACCGCAATGCTGGCCTGCGCTACCGCTCAATGTTGCTGGGCAATGGTCGCTATGTGCCCTTCCA
CATCCAGGTGCCTCAGAAGTTCTTTGCCATTAAAAACCTCCTTCTCCTGCCGGGCTCATACACCTACGAGTGGAACT
TCAGGAAGGATGTTAACATGGTTCTGCAGAGCTCCCTAGGAAATGACCTAAGGGTTGACGGAGCCAGCATTAAGTTT
GATAGCATTTGCCTTTACGCCACCTTCTTCCCCATGGCCCACAACACCGCCTCCACGCTTGAGGCCATGCTTAGAAA
CGACACCAACGACCAGTCCTTTAACGACTATCTCTCCGCCGCCAACATGCTCTACCCTATACCCGCCAACGCTACCA
ACGTGCCCATATCCATCCCCTCCCGCAACTGGGCGGCTTTCCGCGGCTGGGCCTTCACGCGCCTTAAGACTAAGGAA
ACCCCATCACTGGGCTCGGGCTACGACCCTTATTACACCTACTCTGGCTCTATACCCTACCTAGATGGAACCTTTTA
CCTCAACCACACCTTTAAGAAGGTGGCCATTACCTTTGACTCTTCTGTCAGCTGGCCTGGCAATGACCGCCTGCTTA
CCCCCAACGAGTTTGAAATTAAGCGCTCAGTTGACGGGGAGGGTTACAACGTTGCCCAGTGTAACATGACCAAAGAC
TGGTTCCTGGTACAAATGCTAGCTAACTACAACATTGGCTACCAGGGCTTCTATATCCCAGAGAGCTACAAGGACCG
CATGTACTCCTTCTTTAGAAACTTCCAGCCCATGAGCCGTCAGGTGGTGGATGATACTAAATACAAGGACTACCAAC
AGGTGGGCATCCTACACCAACACAACAACTCTGGATTTGTTGGCTACCTTGCCCCCACCATGCGCGAAGGACAGGCC
TACCCTGCTAACTTCCCCTATCCGCTTATAGGCAAGACCGCAGTTGACAGCATTACCCAGAAAAAGTTTCTTTGCGA
TCGCACCCTTTGGCGCATCCCATTCTCCAGTAACTTTATGTCCATGGGCGCACTCACAGACCTGGGCCAAAACCTTC
TCTACGCCAACTCCGCCCACGCGCTAGACATGACTTTTGAGGTGGATCCCATGGACGAGCCCACCCTTCTTTATGTT
TTGTTTGAAGTCTTTGACGTGGTCCGTGTGCACCAGCCGCACCGCGGCGTCATCGAAACCGTGTACCTGCGCACGCC
CTTCTCGGCCGGCAACGCCACAACATAAAGAAGCAAGCAACATCAACAACAGCTGCCGCCATGGGCTCCAGTGAGCA
GGAACTGAAAGCCATTGTCAAAGATCTTGGTTGTGGGCCATATTTTTTGGGCACCTATGACAAGCGCTTTCCAGGCT
TTGTTTCTCCACACAAGCTCGCCTGCGCCATAGTCAATACGGCCGGTCGCGAGACTGGGGGCGTACACTGGATGGCC
TTTGCCTGGAACCCGCACTCAAAAACATGCTACCTCTTTGAGCCCTTTGGCTTTTCTGACCAGCGACTCAAGCAGGT
TTACCAGTTTGAGTACGAGTCACTCCTGCGCCGTAGCGCCATTGCTTCTTCCCCCGACCGCTGTATAACGCTGGAAA
AGTCCACCCAAAGCGTACAGGGGCCCAACTCGGCCGCCTGTGGACTATTCTGCTGCATGTTTCTCCACGCCTTTGCC
AACTGGCCCCAAACTCCCATGGATCACAACCCCACCATGAACCTTATTACCGGGGTACCCAACTCCATGCTCAACAG
TCCCCAGGTACAGCCCACCCTGCGTCGCAACCAGGAACAGCTCTACAGCTTCCTGGAGCGCCACTCGCCCTACTTCC
GCAGCCACAGTGCGCAGATTAGGAGCGCCACTTCTTTTTGTCACTTGAAAAACATGTAAAAATAATGTACTAGAGAC
ACTTTCAATAAAGGCAAATGCTTTTATTTGTACACTCTCGGGTGATTATTTACCCCCACCCTTGCCGTCTGCGCCGT
TTAAAAATCAAAGGGGTTCTGCCGCGCATCGCTATGCGCCACTGGCAGGGACACGTTGCGATACTGGTGTTTAGTGC
TCCACTTAAACTCAGGCACAACCATCCGCGGCAGCTCGGTGAAGTTTTCACTCCACAGGCTGCGCACCATCACCAAC
GCGTTTAGCAGGTCGGGCGCCGATATCTTGAAGTCGCAGTTGGGGCCTCCGCCCTGCGCGCGCGAGTTGCGATACAC
AGGGTTGCAGCACTGGAACACTATCAGCGCCGGGTGGTGCACGCTGGCCAGCACGCTCTTGTCGGAGATCAGATCCG
CGTCCAGGTCCTCCGCGTTGCTCAGGGCGAACGGAGTCAACTTTGGTAGCTGCCTTCCCAAAAAGGGCGCGTGCCCA
GGCTTTGAGTTGCACTCGCACCGTAGTGGCATCAAAAGGTGACCGTGCCCGGTCTGGGCGTTAGGATACAGCGCCTG
CATAAAAGCCTTGATCTGCTTAAAAGCCACCTGAGCCTTTGCGCCTTCAGAGAAGAACATGCCGCAAGACTTGCCGG
AAAACTGATTGGCCGGACAGGCCGCGTCGTGCACGCAGCACCTTGCGTCGGTGTTGGAGATCTGCACCACATTTCGG
CCCCACCGGTTCTTCACGATCTTGGCCTTGCTAGACTGCTCCTTCAGCGCGCGCTGCCCGTTTTCGCTCGTCACATC
CATTTCAATCACGTGCTCCTTATTTATCATAATGCTTCCGTGTAGACACTTAAGCTCGCCTTCGATCTCAGCGCAGC
GGTGCAGCCACAACGCGCAGCCCGTGGGCTCGTGATGCTTGTAGGTCACCTCTGCAAACGACTGCAGGTACGCCTGC
AGGAATCGCCCCATCATCGTCACAAAGGTCTTGTTGCTGGTGAAGGTCAGCTGCAACCCGCGGTGCTCCTCGTTCAG
CCAGGTCTTGCATACGGCCGCCAGAGCTTCCACTTGGTCAGGCAGTAGTTTGAAGTTCGCCTTTAGATCGTTATCCA
CGTGGTACTTGTCCATCAGCGCGCGCGCAGCCTCCATGCCCTTCTCCCACGCAGACACGATCGGCACACTCAGCGGG
TTCATCACCGTAATTTCACTTTCCGCTTCGCTGGGCTCTTCCTCTTCCTCTTGCGTCCGCATACCACGCGCCACTGG
GTCGTCTTCATTCAGCCGCCGCACTGTGCGCTTACCTCCTTTGCCATGCTTGATTAGCACCGGTGGGTTGCTGAAAC
CCACCATTTGTAGCGCCACATCTTCTCTTTCTTCCTCGCTGTCCACGATTACCTCTGGTGATGGCGGGCGCTCGGGC
TTGGGAGAAGGGCGCTTCTTTTTCTTCTTGGGCGCAATGGCCAAATCCGCCGCCGAGGTCGATGGCCGCGGGCTGGG
TGTGCGCGGCACCAGCGCGTCTTGTGATGAGTCTTCCTCGTCCTCGGACTCGATACGCCGCCTCATCCGCTTTTTTG
GGGGCGCCCGGGGAGGCGGCGGCGACGGGGACGGGGACGACACGTCCTCCATGGTTGGGGGACGTCGCGCCGCACCG
CGTCCGCGCTCGGGGGTGGTTTCGCGCTGCTCCTCTTCCCGACTGGCCATTTCCTTCTCCTATAGGCAGAAAAAGAT
CATGGAGTCAGTCGAGAAGAAGGACAGCCTAACCGCCCCCTCTGAGTTCGCCACCACCGCCTCCACCGATGCCGCCA
ACGCGCCTACCACCTTCCCCGTCGAGGCACCCCCGCTTGAGGAGGAGGAAGTGATTATCGAGCAGGACCCAGGTTTT
GTAAGCGAAGACGACGAGGACCGCTCAGTACCAACAGAGGATAAAAAGCAAGACCAGGACAACGCAGAGGCAAACGA
GGAACAAGTCGGGCGGGGGGACGAAAGGCATGGCGACTACCTAGATGTGGGAGACGACGTGCTGTTGAAGCATCTGC
AGCGCCAGTGCGCCATTATCTGCGACGCGTTGCAAGAGCGCAGCGATGTGCCCCTCGCCATAGCGGATGTCAGCCTT
GCCTACGAACGCCACCTATTCTCACCGCGCGTACCCCCCAAACGCCAAGAAAACGGCACATGCGAGCCCAACCCGCG
CCTCAACTTCTACCCCGTATTTGCCGTGCCAGAGGTGCTTGCCACCTATCACATCTTTTTCCAAAACTGCAAGATAC
CCCTATCCTGCCGTGCCAACCGCAGCCGAGCGGACAAGCAGCTGGCCTTGCGGCAGGGCGCTGTCATACCTGATATC

GCCTCGCTCAACGAAGTGCCAAAAATCTTTGAGGGTCTTGGACGCGACGAGAAGCGCGCGGCAAACGCTCTGCAACA
GGAAAACAGCGAAAATGAAAGTCACTCTGGAGTGTTGGTGGAACTCGAGGGTGACAACGCGCGCCTAGCCGTACTAA
AACGCAGCATCGAGGTCACCCACTTTGCCTACCCGGCACTTAACCTACCCCCCAAGGTCATGAGCACAGTCATGAGT
GAGCTGATCGTGCGCCGTGCGCAGCCCCTGGAGAGGGATGCAAATTTGCAAGAACAAACAGAGGAGGGCCTACCCGC
AGTTGGCGACGAGCAGCTAGCGCGCTGGCTTCAAACGCGCGAGCCTGCCGACTTGGAGGAGCGACGCAAACTAATGA
TGGCCGCAGTGCTCGTTACCGTGGAGCTTGAGTGCATGCAGCGGTTCTTTGCTGACCCGGAGATGCAGCGCAAGCTA
GAGGAAACATTGCACTACACCTTTCGACAGGGCTACGTACGCCAGGCCTGCAAGATCTCCAACGTGGAGCTCTGCAA
CCTGGTCTCCTACCTTGGAATTTTGCACGAAAACCGCCTTGGGCAAAACGTGCTTCATTCCACGCTCAAGGGCGAGG
CGCGCCGCGACTACGTCCGCGACTGCGTTTACTTATTTCTATGCTACACCTGGCAGACGGCCATGGGCGTTTGGCAG
CAGTGCTTGGAGGAGTGCAACCTCAAGGAGCTGCAGAAACTGCTAAAGCAAAACTTGAAGGACCTATGGACGGCCTT
CAACGAGCGCTCCGTGGCCGCGCACCTGGCGGACATCATTTTCCCCGAACGCCTGCTTAAAACCCTGCAACAGGGTC
TGCCAGACTTCACCAGTCAAAGCATGTTGCAGAACTTTAGGAACTTTATCCTAGAGCGCTCAGGAATCTTGCCCGCC
ACCTGCTGTGCACTTCCTAGCGACTTTGTGCCCATTAAGTACCGCGAATGCCCTCCGCCGCTTTGGGGCCACTGCTA
CCTTCTGCAGCTAGCCAACTACCTTGCCTACCACTCTGACATAATGGAAGACGTGAGCGGTGACGGTCTACTGGAGT
GTCACTGTCGCTGCAACCTATGCACCCCGCACCGCTCCCTGGTTTGCAATTCGCAGCTGCTTAACGAAAGTCAAATT
ATCGGTACCTTTGAGCTGCAGGGTCCCTCGCCTGACGAAAAGTCCGCGGCTCCGGGGTTGAAACTCACTCCGGGGCT
GTGGACGTCGGCTTACCTTCGCAAATTTGTACCTGAGGACTACCACGCCCACGAGATTAGGTTCTACGAAGACCAAT
CCCGCCCGCCAAATGCGGAGCTTACCGCCTGCGTCATTACCCAGGGCCACATTCTTGGCCAATTGCAAGCCATCAAC
AAAGCCCGCCAAGAGTTTCTGCTACGAAAGGGACGGGGGGTTTACTTGGACCCCCAGTCCGGCGAGGAGCTCAACCC
AATCCCCCCGCCGCCGCAGCCCTATCAGCAGCAGCCGCGGGCCCTTGCTTCCCAGGATGGCACCCAAAAAGAAGCTG
CAGCTGCCGCCGCCACCCACGGACGAGGAGGAATACTGGGACAGTCAGGCAGAGGAGGTTTTGGACGAGGAGGAGGA
GGACATGATGGAAGACTGGGAGAGCCTAGACGAGGAAGCTTCCGAGGTCGAAGAGGTGTCAGACGAAACACCGTCAC
CCTCGGTCGCATTCCCCTCGCCGGCGCCCCAGAAATCGGCAACCGGTTCCAGCATGGCTACAACCTCCGCTCCTCAG
GCGCCGCCGGCACTGCCCGTTCGCCGACCCAACCGTAGATGGGACACCACTGGAACCAGGGCCGGTAAGTCCAAGCA
GCCGCCGCCGTTAGCCCAAGAGCAACAACAGCGCCAAGGCTACCGCTCATGGCGCGGGCACAAGAACGCCATAGTTG
CTTGCTTGCAAGACTGTGGGGGCAACATCTCCTTCGCCCGCCGCTTTCTTCTCTACCATCACGGCGTGGCCTTCCCC
CGTAACATCCTGCATTACTACCGTCATCTCTACAGCCCATACTGCACCGGCGGCAGCGGCAGCGGCAGCAACAGCAG
CGGCCACACAGAAGCAAAGGCGACCGGATAGCAAGACTCTGACAAAGCCCAAGAAATCCACAGCGGCGGCAGCAGCA
GGAGGAGGAGCGCTGCGTCTGGCGCCCAACGAACCCGTATCGACCCGCGAGCTTAGAAACAGGATTTTTCCCACTCT
GTATGCTATATTTCAACAGAGCAGGGGCCAAGAACAAGAGCTGAAAATAAAAAACAGGTCTCTGCGATCCCTCACCC
GCAGCTGCCTGTATCACAAAAGCGAAGATCAGCTTCGGCGCACGCTGGAAGACGCGGAGGCTCTCTTCAGTAAATAC
TGCGCGCTGACTCTTAAGGACTAGTTTCGCGCCCTTTCTCAAATTTAAGCGCGAAAACTACGTCATCTCCAGCGGCC
ACACCCGGCGCCAGCACCTGTCGTCAGCGCCATTATGAGCAAGGAAATTCCCACGCCCTACATGTGGAGTTACCAGC
CACAAATGGGACTTGCGGCTGGAGCTGCCCAAGACTACTCAACCCGAATAAACTACATGAGCGCGGGACCCCACATG
ATATCCCGGGTCAACGGAATCCGCGCCCACCGAAACCGAATTCTCTTGGAACAGGCGGCTATTACCACCACACCTCG
TAATAACCTTAATCCCCGTAGTTGGCCCGCTGCCCTGGTGTACCAGGAAAGTCCCGCTCCCACCACTGTGGTACTTC
CCAGAGACGCCCAGGCCGAAGTTCAGATGACTAACTCAGGGGCGCAGCTTGCGGGCGGCTTTCGTCACAGGGTGCGG
TCGCCCGGGCAGGGTATAACTCACCTGACAATCAGAGGGCGAGGTATTCAGCTCAACGACGAGTCGGTGAGCTCCTC
GCTTGGTCTCCGTCCGGACGGGACATTTCAGATCGGCGGCGCCGGCCGTCCTTCATTCACGCCTCGTCAGGCAATCC
TAACTCTGCAGACCTCGTCCTCTGAGCCGCGCTCTGGAGGCATTGGAACTCTGCAATTTATTGAGGAGTTTGTGCCA
TCGGTCTACTTTAACCCCTTCTCGGGACCTCCCGGCCACTATCCGGATCAATTTATTCCTAACTTTGACGCGGTAAA
GGACTCGGCGGACGGCTACGACTGAATGTTAAGTGGAGAGGCAGAGCAACTGCGCCTGAAACACCTGGTCCACTGTC
GCCGCCACAAGTGCTTTGCCCGCGACTCCGGTGAGTTTTGCTACTTTGAATTGCCCGAGGATCATATCGAGGGCCCG
GCGCACGGCGTCCGGCTTACCGCCCAGGGAGAGCTTGCCCGTAGCCTGATTCGGGAGTTTACCCAGCGCCCCCTGCT
AGTTGAGCGGGACAGGGGACCCTGTGTTCTCACTGTGATTTGCAACTGTCCTAACCTTGGATTACATCAAGATCTTT
GTTGCCATCTCTGTGCTGAGTATAATAAATACAGAAATTAAAATATACTGGGGCTCCTATCGCCATCCTGTAAACGC
CACCGTCTTCACCCGCCCAAGCAAACCAAGGCGAACCTTACCTGGTACTTTTAACATCTCTCCCTCTGTGATTTACA
ACAGTTTCAACCCAGACGGAGTGAGTCTACGAGAGAACCTCTCCGAGCTCAGCTACTCCATCAGAAAAAACACCACC
CTCCTTACCTGCCGGGAACGTACGAGTGCGTCACCGGCCGCTGCACCACACCTACCGCCTGACCGTAAACCAGACTT
TTTCCGGACAGACCTCAATAACTCTGTTTACCAGAACAGGAGGTGAGCTTAGAAAACCCTTAGGGTATTAGGCCAAA
GGCGCAGCTACTGTGGGGTTTATGAACAATTCAAGCAACTCTACGGGCTATTCTAATTCAGGTTTCTCTAGAAATGG
ACGGAATTATTACAGAGCAGCGCCTGCTAGAAAGACGCAGGGCAGCGGCCGAGCAACAGCGCATGAATCAAGAGCTC
CAAGACATGGTTAACTTGCACCAGTGCAAAAGGGGTATCTTTTGTCTGGTAAAGCAGGCCAAAGTCACCTACGACAG
TAATACCACCGGACACCGCCTTAGCTACAAGTTGCCAACCAAGCGTCAGAAATTGGTGGTCATGGTGGGAGAAAAGC
CCATTACCATAACTCAGCACTCGGTAGAAACCGAAGGCTGCATTCACTCACCTTGTCAAGGACCTGAGGATCTCTGC
ACCCTTATTAAGACCCTGTGCGGTCTCAAAGATCTTATTCCCTTTAACTAATAAAAAAAAATAATAAAGCATCACTT
ACTTAAAATCAGTTAGCAAATTTCTGTCCAGTTTATTCAGCAGCACCTCCTTGCCCTCCTCCCAGCTCTGGTATTGC

AGCTTCCTCCTGGCTGCAAACTTTCTCCACAATCTAAATGGAATGTCAGTTTCCTCCTGTTCCTGTCCATCCGCACC
CACTATCTTCATGTTGTTGCAGATGAAGCGCGCAAGACCGTCTGAAGATACCTTCAACCCCGTGTATCCATATGACA
CGGAAACCGGTCCTCCAACTGTGCCTTTTCTTACTCCTCCCTTTGTATCCCCCAATGGGTTTCAAGAGAGTCCCCCT
GGGGTACTCTCTTTGCGCCTATCCGAACCTCTAGTTACCTCCAATGGCATGCTTGCGCTCAAAATGGGCAACGGCCT
CTCTCTGGACGAGGCCGGCAACCTTACCTCCCAAAATGTAACCACTGTGAGCCCACCTCTCAAAAAAACCAAGTCAA
ACATAAACCTGGAAATATCTGCACCCCTCACAGTTACCTCAGAAGCCCTAACTGTGGCTGCCGCCGCACCTCTAATG
GTCGCGGGCAACACACTCACCATGCAATCACAGGCCCCGCTAACCGTGCACGACTCCAAACTTAGCATTGCCACCCA
AGGACCCCTCACAGTGTCAGAAGGAAAGCTAGCCCTGCAAACATCAGGCCCCCTCACCACCACCGATAGCAGTACCC
TTACTATCACTGCCTCACCCCCTCTAACTACTGCCACTGGTAGCTTGGGCATTGACTTGAAAGAGCCCATTTATACA
CAAAATGGAAAACTAGGACTAAAGTACGGGGCTCCTTTGCATGTAACAGACGACCTAAACACTTTGACCGTAGCAAC
TGGTCCAGGTGTGACTATTAATAATACTTCCTTGCAAACTAAAGTTACTGGAGCCTTGGGTTTTGATTCACAAGGCA
ATATGCAACTTAATGTAGCAGGAGGACTAAGGATTGATTCTCAAAACAGACGCCTTATACTTGATGTTAGTTATCCG
TTTGATGCTCAAAACCAACTAAATCTAAGACTAGGACAGGGCCCTCTTTTTATAAACTCAGCCCACAACTTGGATAT
TAACTACAACAAAGGCCTTTACTTGTTTACAGCTTCAAACAATTCCAAAAAGCTTGAGGTTAACCTAAGCACTGCCA
AGGGGTTGATGTTTGACGCTACAGCCATAGCCATTAATGCAGGAGATGGGCTTGAATTTGGTTCACCTAATGCACCA
AACACAAATCCCCTCAAAACAAAAATTGGCCATGGCCTAGAATTTGATTCAAACAAGGCTATGGTTCCTAAACTAGG
AACTGGCCTTAGTTTTGACAGCACAGGTGCCATTACAGTAGGAAACAAAAATAATGATAAGCTAACTTTGTGGACCA
CACCAGCTCCATCTCCTAACTGTAGACTAAATGCAGAGAAAGATGCTAAACTCACTTTGGTCTTAACAAAATGTGGC
AGTCAAATACTTGCTACAGTTTCAGTTTTGGCTGTTAAAGGCAGTTTGGCTCCAATATCTGGAACAGTTCAAAGTGC
TCATCTTATTATAAGATTTGACGAAAATGGAGTGCTACTAAACAATTCCTTCCTGGACCCAGAATATTGGAACTTTA
GAAATGGAGATCTTACTGAAGGCACAGCCTATACAAACGCTGTTGGATTTATGCCTAACCTATCAGCTTATCCAAAA
TCTCACGGTAAAACTGCCAAAAGTAACATTGTCAGTCAAGTTTACTTAAACGGAGACAAAACTAAACCTGTAACACT
AACCATTACACTAAACGGTACACAGGAAACAGGAGACACAACTCCAAGTGCATACTCTATGTCATTTTCATGGGACT
GGTCTGGCCACAACTACATTAATGAAATATTTGCCACATCCTCTTACACTTTTTCATACATTGCCCAAGAATAAAGA
ATCGTTTGTGTTATGTTTCAACGTGTTTATTTTTCAATTGCAGAAAATTTCGAATCATTTTTCATTCAGTAGTATAG
CCCCACCACCACATAGCTTATACAGATCACCGTACCTTAATCAAACTCACAGAACCCTAGTATTCAACCTGCCACCT
CCCTCCCAACACACAGAGTACACAGTCCTTTCTCCCCGGCTGGCCTTAAAAAGCATCATATCATGGGTAACAGACAT
ATTCTTAGGTGTTATATTCCACACGGTTTCCTGTCGAGCCAAACGCTCATCAGTGATATTAATAAACTCCCCGGGCA
GCTCACTTAAGTTCATGTCGCTGTCCAGCTGCTGAGCCACAGGCTGCTGTCCAACTTGCGGTTGCTTAACGGGCGGC
GAAGGAGAAGTCCACGCCTACATGGGGGTAGAGTCATAATCGTGCATCAGGATAGGGCGGTGGTGCTGCAGCAGCGC
GCGAATAAACTGCTGCCGCCGCCGCTCCGTCCTGCAGGAATACAACATGGCAGTGGTCTCCTCAGCGATGATTCGCA
CCGCCCGCAGCATAAGGCGCCTTGTCCTCCGGGCACAGCAGCGCACCCTGATCTCACTTAAATCAGCACAGTAACTG
CAGCACAGCACCACAATATTGTTCAAAATCCCACAGTGCAAGGCGCTGTATCCAAAGCTCATGGCGGGGACCACAGA
ACCCACGTGGCCATCATACCACAAGCGCAGGTAGATTAAGTGGCGACCCCTCATAAACACGCTGGACATAAACATTA
CCTCTTTTGGCATGTTGTAATTCACCACCTCCCGGTACCATATAAACCTCTGATTAAACATGGCGCCATCCACCACC
ATCCTAAACCAGCTGGCCAAAACCTGCCCGCCGGCTATACACTGCAGGGAACCGGGACTGGAACAATGACAGTGGAG
AGCCCAGGACTCGTAACCATGGATCATCATGCTCGTCATGATATCAATGTTGGCACAACACAGGCACACGTGCATAC
ACTTCCTCAGGATTACAAGCTCCTCCCGCGTTAGAACCATATCCCAGGGAACAACCCATTCCTGAATCAGCGTAAAT
CCCACACTGCAGGGAAGACCTCGCACGTAACTCACGTTGTGCATTGTCAAAGTGTTACATTCGGGCAGCAGCGGATG
ATCCTCCAGTATGGTAGCGCGGGTTTCTGTCTCAAAAGGAGGTAGACGATCCCTACTGTACGGAGTGCGCCGAGACA
ACCGAGATCGTGTTGGTCGTAGTGTCATGCCAAATGGAACGCCGGACGTAGTCATATTTCCTGAAGCAAAACCAGGT
GCGGGCGTGACAAACAGATCTGCGTCTCCGGTCTCGCCGCTTAGATCGCTCTGTGTAGTAGTTGTAGTATATCCACT
CTCTCAAAGCATCCAGGCGCCCCCTGGCTTCGGGTTCTATGTAAACTCCTTCATGCGCCGCTGCCCTGATAACATCC
ACCACCGCAGAATAAGCCACACCCAGCCAACCTACACATTCGTTCTGCGAGTCACACACGGGAGGAGCGGGAAGAGC
TGGAAGAACCATGTTTTTTTTTTTATTCCAAAAGATTATCCAAAACCTCAAAATGAAGATCTATTAAGTGAACGCGC
TCCCCTCCGGTGGCGTGGTCAAACTCTACAGCCAAAGAACAGATAATGGCATTTGTAAGATGTTGCACAATGGCTTC
CAAAAGGCAAACGGCCCTCACGTCCAAGTGGACGTAAAGGCTAAACCCTTCAGGGTGAATCTCCTCTATAAACATTC
CAGCACCTTCAACCATGCCCAAATAATTCTCATCTCGCCACCTTCTCAATATATCTCTAAGCAAATCCCGAATATTA
AGTCCGGCCATTGTAAAAATCTGCTCCAGAGCGCCCTCCACCTTCAGCCTCAAGCAGCGAATCATGATTGCAAAAAT
TCAGGTTCCTCACAGACCTGTATAAGATTCAAAAGCGGAACATTAACAAAAATACCGCGATCCCGTAGGTCCCTTCG
CAGGGCCAGCTGAACATAATCGTGCAGGTCTGCACGGACCAGCGCGGCCACTTCCCCGCCAGGAACCTTGACAAAAG
AACCCACACTGATTATGACACGCATACTCGGAGCTATGCTAACCAGCGTAGCCCCGATGTAAGCTTTGTTGCATGGG
CGGCGATATAAAATGCAAGGTGCTGCTCAAAAAATCAGGCAAAGCCTCGCGCAAAAAAGAAAGCACATCGTAGTCAT
GCTCATGCAGATAAAGGCAGGTAAGCTCCGGAACCACCACAGAAAAAGACACCATTTTTCTCTCAAACATGTCTGCG
GGTTTCTGCATAAACACAAAATAAAATAACAAAAAAACATTTAAACATTAGAAGCCTGTCTTACAACAGGAAAAACA
ACCCTTATAAGCATAAGACGGACTACGGCCATGCCGGCGTGACCGTAAAAAAACTGGTCACCGTGATTAAAAAGCAC
CACCGACAGCTCCTCGGTCATGTCCGGAGTCATAATGTAAGACTCGGTAAACACATCAGGTTGATTCACATCGGTCA

GTGCTAAAAAGCGACCGAAATAGCCCGGGGGAATACATACCCGCAGGCGTAGAGACAACATTACAGCCCCCATAGGA
GGTATAACAAAATTAATAGGAGAGAAAAACACATAAACACCTGAAAAACCCTCCTGCCTAGGCAAAATAGCACCCTC
CCGCTCCAGAACAACATACAGCGCTTCCACAGCGGCAGCCATAACAGTCAGCCTTACCAGTAAAAAAGAAAACCTAT
TAAAAAAACACCACTCGACACGGCACCAGCTCAATCAGTCACAGTGTAAAAAAGGGCCAAGTGCAGAGCGAGTATAT
ATAGGACTAAAAAATGACGTAACGGTTAAAGTCCACAAAAAACACCCAGAAAACCGCACGCGAACCTACGCCCAGAA
ACGAAAGCCAAAAAACCCACAACTTCCTCAAATCGTCACTTCCGTTTTCCCACGTTACGTCACTTCCCATTTTAAGA
AAACTACAATTCCCAACACATACAAGTTACTCCGCCCTAAAACCTACGTCACCCGCCCCGTTCCCACGCCCCGCGCC
ACGTCACAAACTCCACCCCCTCATTATCATATTGGCTTCAATCCAAAATAAGGTATATTATTGATGATGTTAATTAA
TTTAAATCCGCATGCGATATCGAGCTCTCCCGGGAATTCGGATCTGCGACGCGAGGCTGGATGGCCTTCCCCATTAT
GATTCTTCTCGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTG
TTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCA
GCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTT
AAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATA
GGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCAC
TCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGC
TCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATA
AAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT
GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATT
TTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTC
ATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGG
TATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGCGATAAGCTCAT
GGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGACAGAACGGTCAGGACCT
GGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGC
CATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCAT
CAGTTCAACGGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGG
AGTCTGATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTG
AGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGAA
AATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTGTC
ATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGT
TGAAGTGGAGCGGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTA
CAGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACT
TATATATTCTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATA
ATTATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAG
GTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCCTTAA
CCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGT
GTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGT
AAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTAT
GCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATG
TTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTC
GCGGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTT
TATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCG
GGTTACAGAACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATAC
GAATCCCTGTGTCAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTA
CCAGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACA
GCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGC
GATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCT
GACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGT
AATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGA
TATCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTG
TACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTC
TTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGTGACTG
AGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGC
GATTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAAT
GAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTA
ATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCT
TCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCA
ACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACG

TATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAG
GATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGG
ATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTT
ATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAG
TTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGC
TCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTG
ATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGAT
CTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGTAATGG
CTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAA
CGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGT
GCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGA
GATTAGATAATGAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATC
GACACCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGC
CTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGG
TCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGG
TCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTG
CTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTG
GGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGG
AAATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGC
CTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTT
ACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGC
TGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATC
AGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACT
GAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCACGTTTTAGTCTA
CGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCA
CTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGT
CTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTG
ATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATC
GTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCC
ACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGAC
CACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGACTATCA
GCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAAC
CTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGT
GGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCACATTTCCCCGAAAAGTGCCACCTGACGTC
TAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGGATCCGA
ATTCCCGGGAGAGCTCGATATCGCATGCGGATTTAAATTAATTAA

B.2 Chapter 4 sequences:

>BAC-Kanstop-Tetstop

ACAATTAATCATCGGCTCGAAGCTTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGG
TGAGGAACTAAACCATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGG
CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCG
CCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTAGC
TGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGC
GAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCG
GCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTC
GGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCC
AGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGT
GGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGG
CTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCC
GATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGGATCAATTCTCTAGAGCTTAATTAAC
GCAGCCTGAATGGCGAATAGGGATCCTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTGCT
AACGCAGTCAGGCACCGTGTATGAATAGTTCGACAAAGATCGCATTGGTAATTACGTTACTCGATGCCATGGGGATT
GGCCTTATCATGCCAGTCTTGCCAACGTTATTACGTGAATTTATTGCTTCGGAAGATATCGCTAACCACTTTGGCGT
ATTGCTTGCACTTTATGCGTTAATGCAGGTTATCTTTGCTCCTTAGCTTGGAAAAATGTCTGACCGATTTGGTCGGC
GCCCAGTGCTGTTGTTGTCATTAATAGGCGCATCGCTGGATTACTTATTGCTGGCTTTTTCAAGTGCGCTTTGGATG
CTGTATTTAGGCCGTTTGCTTTCAGGGATCACAGGAGCTACTGGGGCTGTCGCGGCATCGGTCATTGCCGATACCAC
CTCAGCTTCTCAACGCGTGAAGTGGTTCGGTTGGTTAGGGGCAAGTTTTGGGCTTGGTTTAATAGCGGGGCCTATTA
TTGGTGGTTTTGCAGGAGAGATTTCACCGCATAGTCCCTTTTTTATCGCTGCGTTGCTAAATATTGTCACTTTCCTT
GTGGTTATGTTTTGGTTCCGTGAAACCAAAAATACACGTGATAATACAGATACCGAAGTAGGGGTTGAGACGCAATC
GAATTCGGTATACATCACTTTATTTAAAACGATGCCCATTTTGTTGATTATTTATTTTTCAGCGCAATTGATAGGCC
AAATTCCCGCAACGGTGTGGGTGCTATTTACCGAAAATCGTTTTGGATGGAATAGCATGATGGTTGGCTTTTCATTA
GCGGGTCTTGGTCTTTTACACTCAGTATTCCAAGCCTTTGTGGCAGGAAGAATAGCCACTAAATGGGGCGAAAAAAC
GGCAGTACTGCTCGGATTTATTGCAGATAGTAGTGCATTTGCCTTTTTAGCGTTTATATCTGAAGGTTGGTTAGTTT
TCCCTGTTTTAATTTTATTGGCTGGTGGTGGGATCGCTTTACCTGCATTACAGGGAGTGATGTCTATCCAAACAAAG
AGTCATCAGCAAGGTGCTTTACAGGGATTATTGGTGAGCCTTACCAATGCAACCGGTGTTATTGGCCCATTACTGTT
TGCTGTTATTTATAATCATTCACTACCAATTTGGGATGGCTGGATTTGGATTATTGGTTTAGCGTTTTACTGTATTA
TTATCCTGCTATCGATGACCTTCATGTTAACCCCTCAAGCTCAGGGGAGTAAACAGGAGACAAGTGCTTAGTGATCC
AATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGAC
GTCAGGTGGCACTTTTCGGGGAAATGTGCCCGGTTAACGTGCCGGCACGGCCTGGGTAACCAGGTATTTTGTCCACA
TAACCGTGCGCAAAATGTTGTGGATAAGCAGGACACAGCAGCAATCCACAGCAGGCATACAACCGCACACCGAGGTT
ACTCCGTTCTACAGGTTACGACGACATGTCAATACTTGCCCTTGACAGGCATTGATGGAATCGTAGTCTCACGCTGA
TAGTCTGATCGACAATACAAGTGGGACCGTGGTCCCAGACCGATAATCAGACCGACAACACGAGTGGGATCGTGGTC
CCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGG
ACCGTGGTTCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGAT
ACGAGTGGGACCATGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGTCTGATTATCAG
ACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACT
AATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGTCTGATTATCAGACCGACGATACAAGTGGAACAGTGG
GCCCAGAGAGAATATTCAGGCCAGTTATGCTTTCTGGCCTGTAACAAAGGACATTAAGTAAAGACAGATAAACGTAG
ACTAAAACGTGGTCGCATCAGGGTGCTGGCTTTTCAAGTTCCTTAAGAATGGCCTCAATTTTCTCTATACACTCAGT
TGGAACACGAGACCTGTCCAGGTTAAGCACCATTTTATCGCCCTTATACAATACTGTCGCTCCAGGAGCAAACTGAT
GTCGTGAGCTTAAACTAGTTCTTGATGCAGATGACGTTTTAAGCACAGAAGTTAAAAGAGTGATAACTTCTTCAGCT
TCAAATATCACCCCAGCTTTTTTCTGCTCATGAAGGTTAGATGCCTGCTGCTTAAGTAATTCCTCTTTATCTGTAAA
GGCTTTTTGAAGTGCATCACCTGACCGGGCAGATAGTTCACCGGGGTGAGAAAAAAGAGCAACAACTGATTTAGGCA
ATTTGGCGGTGTTGATACAGCGGGTAATAATCTTACGTGAAATATTTTCCGCATCAGCCAGCGCAGAAATATTTCCA
GCAAATTCATTCTGCAATCGGCTTGCATAACGCTGACCACGTTCATAAGCACTTGTTGGGCGATAATCGTTACCCAA
TCTGGATAATGCAGCCATCTGCTCATCATCCAGCTCGCCAACCAGAACACGATAATCACTTTCGGTAAGTGCAGCAG
CTTTACGACGGCGACTCCCATCGGCAATTTCTATGACACCAGATACTCTTCGACCGAACGCCGGTGTCTGTTGACCA
GTCAGTAGAAAAGAAGGGATGAGATCATCCAGTGCGTCCTCAGTAAGCAGCTCCTGGTCACGTTCATTACCTGACCA
TACCCGAGAGGTCTTCTCAACACTATCACCCCGGAGCACTTCAAGAGTAAACTTCACATCCCGACCACATACAGGCA
AAGTAATGGCATTACCGCGAGCCATTACTCCTACGCGCGCAATTAACGAATCCACCATCGGGGCAGCTGGTGTCGAT
AACGAAGTATCTTCAACCGGTTGAGTATTGAGCGTATGTTTTGGAATAACAGGCGCACGCTTCATTATCTAATCTCC
CAGCGTGGTTTAATCAGACGATCGAAAATTTCATTGCAGACAGGTTCCCAAATAGAAAGAGCATTTCTCCAGGCACC

AGTTGAAGAGCGTTGATCAATGGCCTGTTCAAAAACAGTTCTCATCCGGATCTGACCTTTACCAACTTCATCCGTTT
CACGTACAACATTTTTTAGAACCATGCTTCCCCAGGCATCCCGAATTTGCTCCTCCATCCACGGGGACTGAGAGCCA
TTACTATTGCTGTATTTGGTAAGCAAAATACGTACATCAGGCTCGAACCCTTTAAGATCAACGTTCTTGAGCAGATC
ACGAAGCATATCGAAAAACTGCAGTGCGGAGGTGTAGTCAAACAACTCAGCAGGCGTGGGAACAATCAGCACATCAG
CAGCACATACGACATTAATCGTGCCGATACCCAGGTTAGGCGCGCTGTCAATAACTATGACATCATAGTCATGAGCA
ACAGTTTCAATGGCCAGTCGGAGCATCAGGTGTGGATCGGTGGGCAGTTTACCTTCATCAAATTTGCCCATTAACTC
AGTTTCAATACGGTGCAGAGCCAGACAGGAAGGAATAATGTCAAGCCCCGGCCAGCAAGTGGGCTTTATTGCATAAG
TGACATCGTCCTTTTCCCCAAGATAGAAAGGCAGGAGAGTGTCTTCTGCATGAATATGAAGATCTGGTACCCATCCG
TGATACATTGAGGCTGTTCCCTGGGGGTCGTTACCTTCCACGAGCAAAACACGTAGCCCCTTCAGAGCCAGATCCTG
AGCAAGATGAACAGAAACTGAGGTTTTGTAAACGCCACCTTTATGGGCAGCAACCCCGATCACCGGTGGAAATACGT
CTTCAGCACGTCGCAATCGCGTACCAAACACATCACGCATATGATTAATTTGTTCAATTGTATAACCAACACGTTGC
TCAACCCGTCCTCGAATTTCCATATCCGGGTGCGGTAGTCGCCCTGCTTTCTCGGCATCTCTGATAGCCTGAGAAGA
AACCCCAACTAAATCCGCTGCTTCACCTATTCTCCAGCGCCGGGTTATTTTCCTCGCTTCCGGGCTGTCATCATTAA
ACTGTGCAATGGCGATAGCCTTCGTCATTTCATGACCAGCGTTTATGCACTGGTTAAGTGTTTCCATGAGTTTCATT
CTGAACATCCTTTAATCATTGCTTTGCGTTTTTTTATTAAATCTTGCAATTTACTGCAAAGCAACAACAAAATCGCA
AAGTCATCAAAAAACCGCAAAGTTGTTTAAAATAAGAGCAACACTACAAAAGGAGATAAGAAGAGCACATACCTCAG
TCACTTATTATCACTAGCGCTCGCCGCAGCCGTGTAACCGAGCATAGCGAGCGAACTGGCGAGGAAGCAAAGAAGAA
CTGTTCTGTCAGATAGCTCTTACGCTCAGCGCAAGAAGAAATATCCACCGTGGGAAAAACTCCAGGTAGAGGTACAC
ACGCGGATAGCCAATTCAGAGTAATAAACTGTGATAATCAACCCTCATCAATGATGACGAACTAACCCCCGATATCA
GGTCACATGACGAAGGGAAAGAGAAGGAAATCAACTGTGACAAACTGCCCTCAAATTTGGCTTCCTTAAAAATTACA
GTTCAAAAAGTATGAGAAAATCCATGCAGGCTGAAGGAAACAGCAAAACTGTGACAAATTACCCTCAGTAGGTCAGA
ACAAATGTGACGAACCACCCTCAAATCTGTGACAGATAACCCTCAGACTATCCTGTCGTCATGGAAGTGATATCGCG
GAAGGAAAATACGATATGAGTCGTCTGGCGGCCTTTCTTTTTCTCAATGTATGAGAGGCGCATTGGAGTTCTGCTGT
TGATCTCATTAACACAGACCTGCAGGAAGCGGCGGCGGAAGTCAGGCATACGCTGGTAACTTTGAGGCAGCTGGTAA
CGCTCTATGATCCAGTCGATTTTCAGAGAGACGATGCCTGAGCCATCCGGCTTACGATACTGACACAGGGATTCGTA
TAAACGCATGGCATACGGATTGGTGATTTCTTTTGTTTCACTAAGCCGAAACTGCGTAAACCGGTTCTGTAACCCGA
TAAAGAAGGGAATGAGATATGGGTTGATATGTACACTGTAAAGCCCTCTGGATGGACTGTGCGCACGTTTGATAAAC
CAAGGAAAAGATTCATAGCCTTTTTCATCGCCGGCATCCTCTTCAGGGCGATAAAAAACCACTTCCTTCCCCGCGAA
ACTCTTCAATGCCTGCCGTATATCCTTACTGGCTTCCGCAGAGGTCAATCCGAATATTTCAGCATATTTAGCAACAT
GGATCTCGCAGATACCGTCATGTTCCTGTAGGGTGCCATCAGATTTTCTGATCTGGTCAACGAACAGATACAGCATA
CGTTTTTGATCCCGGGAGAGACTATATGCCGCCTCAGTGAGGTCGTTTGACTGGACGATTCGCGGGCTATTTTTACG
TTTCTTGTGATTGATAACCGCTGTTTCCGCCATGACAGATCCATGTGAAGTGTGACAAGTTTTTAGATTGTCACACT
AAATAAAAAAGAGTCAATAAGCAGGGATAACTTTGTGAAAAAACAGCTTCTTCTGAGGGCAATTTGTCACAGGGTTA
AGGGCAATTTGTCACAGACAGGACTGTCATTTGAGGGTGATTTGTCACACTGAAAGGGCAATTTGTCACAACACCTT
CTCTAGAACCAGCATGGATAAAGGCCTACAAGGCGCTCTAAAAAAGAAGATCTAAAAACTATAAAAAAAATAATTAT
AAAAATATCCCCGTGGATAAGTGGATAACCCCAAGGGAAGTTTTTTCAGGCATCGTGTGTAAGCAGAATATATAAGT
GCTGTTCCCTGGTGCTTCCTCGCTCACTCGAGCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAA
AGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTCATCCATAGTTGCCTGACTGCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTG
CTGCAATGATACCGCGGGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAG
CGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTC
GCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTT
CATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGAAAAAAAGCGGTTAGCTCCTTC
GGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGCTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATC
ATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCG
TGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCG
CAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTAT
CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGGCGGCCGCTTG

>BAC-T7-Kanstop-Tetstop

ACAATTAATCATCGGCTCGAAGCTTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGG
TGAGGAACTAAACCAtGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGG
CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCG

CCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTaGC
TGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGC
GAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCG
GCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTC
GGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCC
AGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGT
GGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGG
CTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCC
GATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGGATCAATTCTCTAGAGCTTAATTAAC
GCAGCCTGAATGGCGAATAGGGATCCTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTGCT
AACGCAGTCAGGCACCGTGTAtGAATAGTTCGACAAAGATCGCATTGGTAATTACGTTACTCGATGCCATGGGGATT
GGCCTTATCATGCCAGTCTTGCCAACGTTATTACGTGAATTTATTGCTTCGGAAGATATCGCTAACCACTTTGGCGT
ATTGCTTGCACTTTATGCGTTAATGCAGGTTATCTTTGCTCCTTaGCTTGGAAAAATGTCTGACCGATTTGGTCGGC
GCCCAGTGCTGTTGTTGTCATTAATAGGCGCATCGCTGGATTACTTATTGCTGGCTTTTTCAAGTGCGCTTTGGATG
CTGTATTTAGGCCGTTTGCTTTCAGGGATCACAGGAGCTACTGGGGCTGTCGCGGCATCGGTCATTGCCGATACCAC
CTCAGCTTCTCAACGCGTGAAGTGGTTCGGTTGGTTAGGGGCAAGTTTTGGGCTTGGTTTAATAGCGGGGCCTATTA
TTGGTGGTTTTGCAGGAGAGATTTCACCGCATAGTCCCTTTTTTATCGCTGCGTTGCTAAATATTGTCACTTTCCTT
GTGGTTATGTTTTGGTTCCGTGAAACCAAAAATACACGTGATAATACAGATACCGAAGTAGGGGTTGAGACGCAATC
GAATTCGGTATACATCACTTTATTTAAAACGATGCCCATTTTGTTGATTATTTATTTTTCAGCGCAATTGATAGGCC
AAATTCCCGCAACGGTGTGGGTGCTATTTACCGAAAATCGTTTTGGATGGAATAGCATGATGGTTGGCTTTTCATTA
GCGGGTCTTGGTCTTTTACACTCAGTATTCCAAGCCTTTGTGGCAGGAAGAATAGCCACTAAATGGGGCGAAAAAAC
GGCAGTACTGCTCGGATTTATTGCAGATAGTAGTGCATTTGCCTTTTTAGCGTTTATATCTGAAGGTTGGTTAGTTT
TCCCTGTTTTAATTTTATTGGCTGGTGGTGGGATCGCTTTACCTGCATTACAGGGAGTGATGTCTATCCAAACAAAG
AGTCATCAGCAAGGTGCTTTACAGGGATTATTGGTGAGCCTTACCAATGCAACCGGTGTTATTGGCCCATTACTGTT
TGCTGTTATTTATAATCATTCACTACCAATTTGGGATGGCTGGATTTGGATTATTGGTTTAGCGTTTTACTGTATTA
TTATCCTGCTATCGATGACCTTCATGTTAACCCCTCAAGCTCAGGGGAGTAAACAGGAGACAAGTGCTTAGTGATCC
AATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGAC
GTCAGGTGGCACTTTTCGGGGAAATGTGCCCGGTTAACGTGCCGGCACGGCCTGGGTAACCAGGTATTTTGTCCACA
TAACCGTGCGCAAAATGTTGTGGATAAGCAGGACACAGCAGCAATCCACAGCAGGCATACAACCGCACACCGAGGTT
ACTCCGTTCTACAGGTTACGACGACATGTCAATACTTGCCCTTGACAGGCATTGATGGAATCGTAGTCTCACGCTGA
TAGTCTGATCGACAATACAAGTGGGACCGTGGTCCCAGACCGATAATCAGACCGACAACACGAGTGGGATCGTGGTC
CCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGG
ACCGTGGTTCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGAT
ACGAGTGGGACCATGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGTCTGATTATCAG
ACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACT
AATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGTCTGATTATCAGACCGACGATACAAGTGGAACAGTGG
GCCCAGAGAGAATATTCAGGCCAGTTATGCTTTCTGGCCTGTAACAAAGGACATTAAGTAAAGACAGATAAACGTAG
ACTAAAACGTGGTCGCATCAGGGTGCTGGCTTTTCAAGTTCCTTAAGAATGGCCTCAATTTTCTCTATACACTCAGT
TGGAACACGAGACCTGTCCAGGTTAAGCACCATTTTATCGCCCTTATACAATACTGTCGCTCCAGGAGCAAACTGAT
GTCGTGAGCTTAAACTAGTTCTTGATGCAGATGACGTTTTAAGCACAGAAGTTAAAAGAGTGATAACTTCTTCAGCT
TCAAATATCACCCCAGCTTTTTTCTGCTCATGAAGGTTAGATGCCTGCTGCTTAAGTAATTCCTCTTTATCTGTAAA
GGCTTTTTGAAGTGCATCACCTGACCGGGCAGATAGTTCACCGGGGTGAGAAAAAAGAGCAACAACTGATTTAGGCA
ATTTGGCGGTGTTGATACAGCGGGTAATAATCTTACGTGAAATATTTTCCGCATCAGCCAGCGCAGAAATATTTCCA
GCAAATTCATTCTGCAATCGGCTTGCATAACGCTGACCACGTTCATAAGCACTTGTTGGGCGATAATCGTTACCCAA
TCTGGATAATGCAGCCATCTGCTCATCATCCAGCTCGCCAACCAGAACACGATAATCACTTTCGGTAAGTGCAGCAG
CTTTACGACGGCGACTCCCATCGGCAATTTCTATGACACCAGATACTCTTCGACCGAACGCCGGTGTCTGTTGACCA
GTCAGTAGAAAAGAAGGGATGAGATCATCCAGTGCGTCCTCAGTAAGCAGCTCCTGGTCACGTTCATTACCTGACCA
TACCCGAGAGGTCTTCTCAACACTATCACCCCGGAGCACTTCAAGAGTAAACTTCACATCCCGACCACATACAGGCA
AAGTAATGGCATTACCGCGAGCCATTACTCCTACGCGCGCAATTAACGAATCCACCATCGGGGCAGCTGGTGTCGAT
AACGAAGTATCTTCAACCGGTTGAGTATTGAGCGTATGTTTTGGAATAACAGGCGCACGCTTCATTATCTAATCTCC
CAGCGTGGTTTAATCAGACGATCGAAAATTTCATTGCAGACAGGTTCCCAAATAGAAAGAGCATTTCTCCAGGCACC
AGTTGAAGAGCGTTGATCAATGGCCTGTTCAAAAACAGTTCTCATCCGGATCTGACCTTTACCAACTTCATCCGTTT
CACGTACAACATTTTTTAGAACCATGCTTCCCCAGGCATCCCGAATTTGCTCCTCCATCCACGGGGACTGAGAGCCA
TTACTATTGCTGTATTTGGTAAGCAAAATACGTACATCAGGCTCGAACCCTTTAAGATCAACGTTCTTGAGCAGATC
ACGAAGCATATCGAAAAACTGCAGTGCGGAGGTGTAGTCAAACAACTCAGCAGGCGTGGGAACAATCAGCACATCAG
CAGCACATACGACATTAATCGTGCCGATACCCAGGTTAGGCGCGCTGTCAATAACTATGACATCATAGTCATGAGCA
ACAGTTTCAATGGCCAGTCGGAGCATCAGGTGTGGATCGGTGGGCAGTTTACCTTCATCAAATTTGCCCATTAACTC

AGTTTCAATACGGTGCAGAGCCAGACAGGAAGGAATAATGTCAAGCCCCGGCCAGCAAGTGGGCTTTATTGCATAAG
TGACATCGTCCTTTTCCCCAAGATAGAAAGGCAGGAGAGTGTCTTCTGCATGAATATGAAGATCTGGTACCCATCCG
TGATACATTGAGGCTGTTCCCTGGGGGTCGTTACCTTCCACGAGCAAAACACGTAGCCCCTTCAGAGCCAGATCCTG
AGCAAGATGAACAGAAACTGAGGTTTTGTAAACGCCACCTTTATGGGCAGCAACCCCGATCACCGGTGGAAATACGT
CTTCAGCACGTCGCAATCGCGTACCAAACACATCACGCATATGATTAATTTGTTCAATTGTATAACCAACACGTTGC
TCAACCCGTCCTCGAATTTCCATATCCGGGTGCGGTAGTCGCCCTGCTTTCTCGGCATCTCTGATAGCCTGAGAAGA
AACCCCAACTAAATCCGCTGCTTCACCTATTCTCCAGCGCCGGGTTATTTTCCTCGCTTCCGGGCTGTCATCATTAA
ACTGTGCAATGGCGATAGCCTTCGTCATTTCATGACCAGCGTTTATGCACTGGTTAAGTGTTTCCATGAGTTTCATT
CTGAACATCCTTTAATCATTGCTTTGCGTTTTTTTATTAAATCTTGCAATTTACTGCAAAGCAACAACAAAATCGCA
AAGTCATCAAAAAACCGCAAAGTTGTTTAAAATAAGAGCAACACTACAAAAGGAGATAAGAAGAGCACATACCTCAG
TCACTTATTATCACTAGCGCTCGCCGCAGCCGTGTAACCGAGCATAGCGAGCGAACTGGCGAGGAAGCAAAGAAGAA
CTGTTCTGTCAGATAGCTCTTACGCTCAGCGCAAGAAGAAATATCCACCGTGGGAAAAACTCCAGGTAGAGGTACAC
ACGCGGATAGCCAATTCAGAGTAATAAACTGTGATAATCAACCCTCATCAATGATGACGAACTAACCCCCGATATCA
GGTCACATGACGAAGGGAAAGAGAAGGAAATCAACTGTGACAAACTGCCCTCAAATTTGGCTTCCTTAAAAATTACA
GTTCAAAAAGTATGAGAAAATCCATGCAGGCTGAAGGAAACAGCAAAACTGTGACAAATTACCCTCAGTAGGTCAGA
ACAAATGTGACGAACCACCCTCAAATCTGTGACAGATAACCCTCAGACTATCCTGTCGTCATGGAAGTGATATCGCG
GAAGGAAAATACGATATGAGTCGTCTGGCGGCCTTTCTTTTTCTCAATGTATGAGAGGCGCATTGGAGTTCTGCTGT
TGATCTCATTAACACAGACCTGCAGGAAGCGGCGGCGGAAGTCAGGCATACGCTGGTAACTTTGAGGCAGCTGGTAA
CGCTCTATGATCCAGTCGATTTTCAGAGAGACGATGCCTGAGCCATCCGGCTTACGATACTGACACAGGGATTCGTA
TAAACGCATGGCATACGGATTGGTGATTTCTTTTGTTTCACTAAGCCGAAACTGCGTAAACCGGTTCTGTAACCCGA
TAAAGAAGGGAATGAGATATGGGTTGATATGTACACTGTAAAGCCCTCTGGATGGACTGTGCGCACGTTTGATAAAC
CAAGGAAAAGATTCATAGCCTTTTTCATCGCCGGCATCCTCTTCAGGGCGATAAAAAACCACTTCCTTCCCCGCGAA
ACTCTTCAATGCCTGCCGTATATCCTTACTGGCTTCCGCAGAGGTCAATCCGAATATTTCAGCATATTTAGCAACAT
GGATCTCGCAGATACCGTCATGTTCCTGTAGGGTGCCATCAGATTTTCTGATCTGGTCAACGAACAGATACAGCATA
CGTTTTTGATCCCGGGAGAGACTATATGCCGCCTCAGTGAGGTCGTTTGACTGGACGATTCGCGGGCTATTTTTACG
TTTCTTGTGATTGATAACCGCTGTTTCCGCCATGACAGATCCATGTGAAGTGTGACAAGTTTTTAGATTGTCACACT
AAATAAAAAAGAGTCAATAAGCAGGGATAACTTTGTGAAAAAACAGCTTCTTCTGAGGGCAATTTGTCACAGGGTTA
AGGGCAATTTGTCACAGACAGGACTGTCATTTGAGGGTGATTTGTCACACTGAAAGGGCAATTTGTCACAACACCTT
CTCTAGAACCAGCATGGATAAAGGCCTACAAGGCGCTCTAAAAAAGAAGATCTAAAAACTATAAAAAAAATAATTAT
AAAAATATCCCCGTGGATAAGTGGATAACCCCAAGGGAAGTTTTTTCAGGCATCGTGTGTAAGCAGAATATATAAGT
GCTGTTCCCTGGTGCTTCCTCGCTCACTCGAGCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAA
AGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTCATCCATAGTTGCCTGACTGCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTG
CTGCAATGATACCGCGGGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAG
CGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTC
GCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTT
CATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGAAAAAAAGCGGTTAGCTCCTTC
GGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGCTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATC
ATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCG
TGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCG
CAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTAT
CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGGCGGCCGCCTTTCAGCAAAAAACCCCGCGAGACCCCCGAAGAGGCCCCGCGGGGTTA
TGCTAGGTCGACGGAGCTCGAATTCTAATACGACTCACTATAGGGAGACCCAAGCTGGCTTG

B.3 Chapter 5 sequences:

>pBAD-recA

GCTCCGGGCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGGAC
TAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACAT
AAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACGGGCGAAACCCGAC
AGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGG
ATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTG
TAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTG
TTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGG
ATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTG
CTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTAACCTTCGATCAAACCACCTCCCCA
GGTGGTTTTTTTCGTTTACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
CTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCAGGAGGAAGAGGACAT
CCGGTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAGACTTAGGGACCCTTTATGACAACTT
GACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATA
CCCGCGAGAAATAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAA
AGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATG
TGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGC
TGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGC
AGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTG
CCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCC
AGTTAAGCCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGA
CGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAATTCTCGTCCCTGATT
TTTCACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAAT
CGAGATAACCGTTGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGA
TCATTTTGCGCTTCAGCCATACTTTTCATACTCCCACCATTCAGAGAAGAAACCAATTGTCCATATTGCATCAGACA
TTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCCAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAAC
AAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGATTAT
TTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATC
GCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGACGCGTACAACTCAAGGAGGTGGTACCATGGCTATCGACG
AAAACAAACAGAAAGCGTTGGCGGCAGCACTGGGCCAGATTGAGAAACAATTTGGTAAAGGCTCCATCATGCGCCTG
GGTGAAGACCGTTCCATGGATGTGGAAACCATCTCTACCGGTTCGCTTTCACTGGATATCGCGCTTGGGGCAGGTGG
TCTGCCGATGGGCCGTATCGTCGAAATCTACGGACCGGAATCTTCCGGTAAAACCACGCTGACGCTGCAGGTGATCG
CCGCAGCGCAGCGTGAAGGTAAAACCTGTGCGTTTATCGATGCTGAACACGCGCTGGACCCAATCTACGCACGTAAA
CTGGGCGTCGATATCGACAACCTGCTGTGCTCCCAGCCGGACACCGGCGAGCAGGCACTGGAAATCTGTGACGCCCT
GGCGCGTTCTGGCGCAGTAGACGTTATCGTCGTTGACTCCGTGGCGGCACTGACGCCGAAAGCGGAAATCGAAGGCG
AAATCGGCGACTCTCACATGGGCCTTGCGGCACGTATGATGAGCCAGGCGATGCGTAAGCTGGCGGGTAACCTGAAG
CAGTCCAACACGCTGCTGATCTTCATCAACCAGATCCGTATGAAAATTGGTGTGATGTTCGGTAACCCGGAAACCAC
TACCGGTGGTAACGCGCTGAAATTCTACGCCTCTGTTCGTCTCGACATCCGTCGTATCGGCGCGGTGAAAGAGGGCG
AAAACGTGGTGGGTAGCGAAACCCGCGTGAAAGTGGTGAAGAACAAAATCGCTGCGCCGTTTAAACAGGCTGAATTC
CAGATCCTCTACGGCGAAGGTATCAACTTCTACGGCGAACTGGTTGACCTGGGCGTAAAAGAGAAGCTGATCGAGAA
AGCAGGCGCGTGGTACAGCTACAAAGGTGAGAAGATCGGTCAGGGTAAAGCGAATGCGACTGCCTGGCTGAAAGATA
ACCCGGAAACCGCGAAAGAGATCGAGAAGAAAGTACGTGAGTTGCTGCTGAGCAACCCGAACTCAACGCCGGATTTC
TCTGTAGATGATAGCGAAGGCGTAGCAGAAACTAACGAAGATTTTTAAGCTCTCGAGACTTAATTAACGGCACTCCT
CAGCCAAGTCAAAAGCCTCCGGTCGGAGGCTTTTGACTACATGGCGGCCGCGTCGAGGCTGATCAGCGAGCTCTAGA
GAATTGATCCCCTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGT
AAAGCACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGA
TAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAA
GCAGGCATCGCCATGGGTCACGACGAGATCATCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGAACAGTTCGGCTG
GCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCG
ATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGCATTGCATCAGCCAT
GATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGT
CCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCT
GCCTCGTCCTGCAGTTCATTCAGGGCACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCCTGCGCTGACAG
CCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGG
CCGGAGAACCTGCGTGCAATCCATCTTGTTCAATGGCCGATCCCATGGTTTAGTTCCTCACCTTGTCGTATTATACT

ATGCCGATATACTATGCCGATGATTAATTGTCAACAGGCTGCAGGTCGAAAGAATTCTTCATTATGGTGAAAGTTGG
AACCTCTTACGTGCCAAGCCAAATAGGCCGTCACTCGGTCGCTAC

>pL-lacZ

AGCTTGATCTCTCACCTACCAAACAATGCCCCCCTGCAAAAAATAAATTCATATAAAAAACATACAGATAACCATCT
GCGGTGATAAATTATCTCTGGCGGTGTTGACATAAATACCACTGGCGGTGATACTGAGCACATCAGCAGGACGCACT
GACCACCATGAAGGTGACGCTCTTAAAAATTAAGCCCTGAAGAAGGGCCTCGAGGCGGCCGCAGGAGGTACAGCTAT
GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
GCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTG
CGCAGCCTGAATGGCGAATAGGGATCCTCTAGACCGGTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAA
CCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAA
TAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCT
TCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACA
TCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTT
AAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTC
TCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCA
GTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACC
GCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTC
TAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCG
GCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGA
TGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCG
CTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTA
AAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGA
GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA
TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTT
CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT
CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGC
ACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCT
TCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG
GGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG
TCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTC
GCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTC
CCCGCGCGTTGGCCGATTCATTAATGCAGA

>BAC-cl

CCTGACGTCTAAGAAACCATTATTATCATGACATTGAACAATTAATCATCGGCTCGTATAATGTGTGGAAACAGCGG
TGAAGGAGGTAACGTATGAGCACAAAAAAGAAACCATTAACACAAGAGCAGCTTGAGGACGCACGTCGCCTTAAAGC
AATTTATGAAAAAAAGAAAAATGAACTTGGCTTATCCCAGGAATCTGTCGCAGACAAGATGGGGATGGGGCAGTCAG
GCGTTGGTGCTTTATTTAATGGCATCAATGCATTAAATGCTTATAACGCCGCATTGCTTACAAAAATTCTCAAAGTT
AGCGTTGAAGAATTTAGCCCTTCAATCGCCAGAGAAATCTACGAGATGTATGAAGCGGTTAGTATGCAGCCGTCACT
TAGAAGTGAGTATGAGTACCCTGTTTTTTCTCATGTTCAGGCAGGGATGTTCTCACCTGAGCTTAGAACCTTTACCA
AAGGTGATGCGGAGAGATGGGTAAGCACAACCAAAAAAGCCAGTGATTCTGCATTCTGGCTTGAGGTTGAAGGTAAT
TCCATGACCGCACCAACAGGCTCCAAGCCAAGCTTTCCTGACGGAATGTTAATTCTCGTTGACCCTGAGCAGGCTGT
TGAGCCAGGTGATTTCTGCATAGCCAGACTTGGGGGTGATGAGTTTACCTTCAAGAAACTGATCAGGGATAGCGGTC
AGGTGTTTTTACAACCACTAAACCCACAGTACCCAATGATCCCATGCAATGAGAGTTGTTCCGTTGTGGGGAAAGTT
ATCGCTAGTCAGTGGCCTGAAGAGACGTTTGGCTGACGGATCTGCGACGCGAGGCTGGATGGCCTTCCCCATTATGA
TTCTTCTCGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTT
GTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGC
ACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAA
ATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGG
CCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTC
CAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTC
ACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAA
ACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGA
GCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTT

TTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCAT
TATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTA
TCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGCGATAAGCTCATGG
AGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGACAGAACGGTCAGGACCTGG
ATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGCCA
TTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCA
GTTCAACGGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAG
TCTGATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAG
TGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGAAAA
TCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTGTCAT
GATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGTTG
AAGTGGAGCGGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACA
GCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTA
TATATTCTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAAT
TATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGT
GTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCCTTAACC
CTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGTGT
GACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGTAA
AAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATGC
TGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTT
GCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGC
GGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTA
TCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGG
TTACAGAACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGA
ATCCCTGTGTCAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACC
AGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGC
AGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGCGA
TATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGA
CCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAA
TTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATA
TCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTA
CCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTT
CTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGTGACTGAG
GTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGA
TTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATGA
AACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAAT
GATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTC
TCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCAAC
GTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTA
TTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGA
TCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGAT
GGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTAT
GCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTT
AATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTC
ATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGAT
GTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCT
GCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGTAATGGCT
CTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACG
GATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGC
CTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGA
TTAGATAATGAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGA
CACCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCT
GTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTC
AGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTC
AACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTGCT
GCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGG
TAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAA

ATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCT
AAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTAC
AGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGCTG
AAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAG
TTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGA
GTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCACGTTTTAGTCTACG
TTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACT
GTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCT
GGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGAT
AATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGT
CGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCAC
TCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCA
CGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGC
GTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACCT
CGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGG
ACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCACATTTCCCCGAAAAGTGCCA

>CY2115

GTTGCTCCGGGCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAAGTTGGCATTATAAAAAAG
CATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGATTTCAATTTTGTCC
CACTCCCTGCCTCTGTCATCACGATACTGTGATGCCATGGTGTCCGACTTATGCCCGAGAAGATGTTGAGCAAACTT
ATCGCTTATCTGCTTCTCATAGAGTCTTGCAGACAAACTGCGCAACTCGTGAAAGGTAGGCGGATCCCCTTCGAAGG
AAAGACCTGATGCTTTTCGTGCGCGCATAAAATACCTTGATACTGTGCCGGATGAAAGCGGTTCGCGACGAGTAGAT
GCAATTATGGTTTCTCCGCCAAGAATCTCTTTGCATTTATCAAGTGTTTCCTTCATTGATATTCCGAGAGCATCAAT
ATGCAATGCTGTTGGGATGGCAATTTTTACGCCTGTTTTGCTTTGCTCGACATAAAGATATCCATCTACGATATCAG
ACCACTTCATTTCGCATAAATCACCAACTCGTTGCCCGGTAACAACAGCCAGTTCCATTGCAAGTCTGAGCCAACAT
GGTGATGATTCTGCTGCTTGATAAATTTTCAGGTATTCGTCAGCCGTAAGTCTTGATCTCCTTACCTCTGATTTTGC
TGCGCGAGTGGCAGCGACATGGTTTGTTGTTATATGGCCTTCAGCTATTGCCTCTCGGAATGCATCGCTCAGTGTTG
ATCTGATTAACTTGGCTGACGCCGCCTTGCCCTCGTCTATGTATCCATTGAGCATTGCCGCAATTTCTTTTGTGGTG
ATGTCTTCAAGTGGAGCATCAGGCAGACCCCTCCTTATTGCTTTAATTTTGCTCATGTAATTTATGAGTGTCTTCTG
CTTGATTCCTCTGCTGGCCAGGATTTTTTCGTAGCGATCAAGCCATGAATGTAACGTAACGGAATTATCACTGTTGA
TTCTCGCTGTCAGAGGCTTGTGTTTGTGTCCTGAAAATAACTCAATGTTGGCCTGTATAGCTTCAGTGATTGCGATT
CGCCTGTCTCTGCCTAATCCAAACTCTTTACCCGTCCTTGGGTCCCTGTAGCAGTAATATCCATTGTTTCTTATATA
AAGGTTAGGGGGTAAATCCCGGCGCTCATGACTTCGCCTTCTTCCCATTTCTGATCCTCTTCAAAAGGCCACCTGTT
ACTGGTCGATTTAAGTCAACCTTTACCGCTGATTCGTGGAACAGATACTCTCTTCCATCCTTAACCGGAGGTGGGAA
TATCCTGCATTCCCGAACCCATCGACGAACTGTTTCAAGGCTTCTTGGACGTCGCTGGCGTGCGTTCCACTCCTGAA
GTGTCAAGTACATCGCAAAGTCTCCGCAATTACACGCAAGAAAAAACCGCCATCAGGCGGCTTGGTGTTCTTTCAGT
TCTTCAATTCGAATATTGGTTACGTCTGCATGTGCTATCTGCGCCCATATCATCCAGTGGTCGTAGCAGTCGTTGAT
GTTCTCCGCTTCGATAACTCTGTTGAATGGCTCTCCATTCCATTCTCCTGTGACTCGGAAGTGCATTTATCATCTCC
ATAAAACAAAACCCGCCGTAGCGAGTTCAGATAAAATAAATCCCCGCGAGTGCGAGGATTGTTATGTAATATTGGGT
TTAATCATCTATATGTTTTGTACAGAGAGGGCAAGTATCGTTTCCACCGTACTCGTGATAATAATTTTGCACGGTAT
CAGTCATTTCTCGCACATTGCAGAATGGGGATTTGTCTTCATTAGACTTATAAACCTTCATGGAATATTTGTATGCC
GACTCTATATCTATACCTTCATCTACATAAACACCTTCGTGATGTCTGCATGGAGACAAGACACCGGATCTGCACAA
CATTGATAACGCCCAATCTTTTTGCTCAGACTCTAACTCATTGATACTCATTTATAAACTCCTTGCAATGTATGTCG
TTTCAGCTAAACGGTATCAGCAATGTTTATGTAAAGAAACAGTAAGATAATACTCAACCCGATGTTTGAGTACGGTC
ATCATCTGACACTACAGACTCTGGCATCGCTGTGAAGACGACGCGAAATTCAGCATTTTCACAAGCGTTATCTTTTA
CAAAACCGATCTCACTCTCCTTTGATGCGAATGCCAGCGTCAGACATCATATGCAGATACTCACCTGCATCCTGAAC
CCATTGACCTCCAACCCCGTAATAGCGATGCGTAATGATGTCGATAGTTACTAACGGGTCTTGTTCGATTAACTGCC
GCAGAAACTCTTCCAGGTCACCAGTGCAGTGCTTGATAACAGGAGTCTTCCCAGGATGGCGAACAACAAGAAACTGG
TTTCCGTCTTCACGGACTTCGTTGCTTTCCAGTTTAGCAATACGCTTACTCCCATCCGAGATAACACCTTCGTAATA
CTCACGCTGCTCGTTGAGTTTTGATTTTGCTGTTTCAAGCTCAACACGCAGTTTCCCTACTGTTAGCGCAATATCCT
CGTTCTCCTGGTCGCGGCGTTTGATGTATTGCTGGTTTCTTTCCCGTTCATCCAGCAGTTCCAGCACAATCGATGGT
GTTACCAATTCATGGAAAAGGTCTGCGTCAAATCCCCAGTCGTCATGCATTGCCTGCTCTGCCGCTTCACGCAGTGC
CTGAGAGTTAATTTCGCTCACTTCGAACCTCTCTGTTTACTGATAAGTTCCAGATCCTCCTGGCAACTTGCACAAGT
CCGACAACCCTGAACGACCAGGCGTCTTCGTTCATCTATCGGATCGCCACACTCACAACAATGAGTGGCAGATATAG
CCTGGTGGTTCAGGCGGCGCATTTTTATTGCTGTGTTGCGCTGTAATTCTTCTATTTCTGATGCTGAATCAATGATG
TCTGCCATCTTTCATTAATCCCTGAACTGTTGGTTAATACGCTTGAGGGTGAATGCGAATAATAAAAAAGGAGCCTG

TAGCTCCCTGATGATTTTGCTTTTCATGTTCATCGTTCCTTAAAGACGCCGTTTAACATGCCGATTGCCAGGCTTAA
ATGAGTCGGTGTGAATCCCATCAGCGTTACCGTTTCGCGGTGCTTCTTCAGTACGCTACGGCAAATGTCATCGACGT
TTTTATCCGGAAACTGCTGTCTGGCTTTTTTTGATTTCAGAATTAGCCTGACGGGCAATGCTGCGAAGGGCGTTTTC
CTGCTGAGGTGTCATTGAACAAGTCCCATGTCGGCAAGCATAAGCACACAGAATATGAAGCCCGCTGCCAGAAAAAT
GCATTCCGTGGTTGTCATACCTGGTTTCTCTCATCTGCTTCTGCTTTCGCCACCATCATTTCCAGCTTTTGTGAAAG
GGATGCGGCTAACGTATGAAATTCTTCGTCTGTTTCTACTGGTATTGGCACAAACCTGATTCCAATTTGAGCAAGGC
TATGTGCCATCTCGATACTCGTTCTTAACTCAACAGAAGATGCTTTGTGCATACAGCCCCTCGTTTATTATTTATCT
CCTCAGCCAGCCGCTGTGCTTTCAGTGGATTTCGGATAACAGAAAGGCCGGGAAATACCCAGCCTCGCTTTGTAACG
GAGTAGACGAAAGTGATTGCGCCTACCCGGATATTATCGTGAGGATGCGTCATCGCCATTGCTCCCCAAATACAAAA
CCAATTTCAGCCAGTGCCTCGTCCATTTTTTCGATGAACTCCGGCACGATCTCGTCAAAACTCGCCATGTACTTTTC
ATCCCGCTCAATCACGACATAATGCAGGCCTTCACGCTTCATACGCGGGTCATAGTTGGCAAAGTACCAGGCATTTT
TTCGCGTCACCCACATGCTGTACTGCACCTGGGCCATGTAAGCTGACTTTATGGCCTCGAAACCACCGAGCCGGAAC
TTCATGAAATCCCGGGAGGTAAACGGGCATTTCAGTTCAAGGCCGTTGCCGTCACTGCATAAACCATCGGGAGAGCA
GGCGGTACGCATACTTTCGTCGCGATAGATGATCGGGGATTCAGTAACATTCACGCCGGAAGTGAATTCAAACAGGG
TTCTGGCGTCGTTCTCGTACTGTTTTCCCCAGGCCAGTGCTTTAGCGTTAACTTCCGGAGCCACACCGGTGCAAACC
TCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATGTCAGGCCACTTCTTTCCGGAGCGGGGTTTTGCTATCACGTT
GTGAACTTCTGAAGCGGTGATGACGCCGAGCCGTAATTTGTGCCACGCATCATCCCCCTGTTCGACAGCTCTCACAT
CGATCCCGGTACGCTGCAGGATAATGTCCGGTGTCATGCTGCCACCTTCTGCTCTGCGGCTTTCTGTTTCAGGAATC
CAAGAGCTTTTACTGCTTCGGCCTGTGTCAGTTCTGACGATGCACGAATGTCGCGGCGAAATATCTGGGAACAGAGC
GGCAATAAGTCGTCATCCCATGTTTTATCCAGGGCGATCAGCAGAGTGTTAATCTCCTGCATGGTTTCATCGTTAAC
CGGAGTGATGTCGCGTTCCGGCTGACGTTCTGCAGTGTATGCAGTATTTTCGACAATGCGCTCGGCTTCATCCTTGT
CATAGATACCAGCAAATCCGAAGGCCAGACGGGCACACTGAATCATGGCTTTATGACGTAACATCCGTTTGGGATGC
GACTGCCACGGCCCCGTGATTTCTCTGCCTTCGCGAGTTTTGAATGGTTCGCGGCGGCATTCATCCATCCATTCGGT
AACGCAGATCGGATGATTACGGTCCTTGCGGTAAATCCGGCATGTACAGGATTCATTGTCCTGCTCAAAGTCCATGC
CATCAAACTGCTGGTTTTCATTGATGATGCGGGACCAGCCATCAACGCCCACCACCGGAACGATGCCATTCTGCTTA
TCAGGAAAGGCGTAAATTTCTTTCGTCCACGGATTAAGGCCGTACTGGTTGGCAACGATCAGTAATGCGATGAACTG
CGCATCGCTGGCATCACCTTTAAATGCCGTCTGGCGAAGAGTGGTGATCAGTTCCTGTGGGTCGACAGAATCCATGC
CGACACGTTCAGCCAGCTTCCCAGCCAGCGTTGCGAGTGCAGTACTCATTCGTTTTATACCTCTGAATCAATATCAA
CCTGGTGGTGAGCAATGGTTTCAACCATGTACCGGATGTGTTCTGCCATGCGCTCCTGAAACTCAACATCGTCATCA
AACGCACGGGTAATGGATTTTTTGCTGGCCCCGTGGCGTTGCAAATGATCGATGCATAGCGATTCAAACAGGTGCTG
GGGCAGGCCTTTTTCCATGTCGTCTGCCAGTTCTGCCTCTTTCTCTTCACGGGCGAGCTGCTGGTAGTGACGCGCCC
AGCTCTGAGCCTCAAGACGATCCTGAATGTAATAAGCGTTCATGGCTGAACTCCTGAAATAGCTGTGAAAATATCGC
CCGCGAAATGCCGGGCTGATTAGGAAAACAGGAAAGGGGGTTAGTGAATGCTTTTGCTTGATCTCAGTTTCAGTATT
AATATCCATTTTTTATAAGCGTCGACGGCTTCACGAAACATCTTTTCATCGCCAATAAAAGTGGCGATAGTGAATTT
AGTCTGGATAGCCATAAGTGTTTGATCCATTCTTTGGGACTCCTGGCTGATTAAGTATGTCGATAAGGCGTTTCCAT
CCGTCACGTAATTTACGGGTGATTCGTTCAAGTAAAGATTCGGAAGGGCAGCCAGCAACAGGCCACCCTGCAATGGC
ATATTGCATGGTGTGCTCCTTATTTATACATAACGAAAAACGCCTCGAGTGAAGCGTTATTGGTATGCGGTAAAACC
GCACTCAGGCGGCCTTGATAGTCATATCATCTGAATCAAATATTCCTGATGTATCGATATCGGTAATTCTTATTCCT
TCGCTACCATCCATTGGAGGCCATCCTTCCTGACCATTTCCATCATTCCAGTCGAACTCACACACAACACCATATGC
ATTTAAGTCGCTTGAAATTGCTATAAGCAGAGCATGTTGCGCCAGCATGATTAATACAGCATTTAATACAGAGCCGT
GTTTATTGAGTCGGTATTCAGAGTCTGACCAGAAATTATTAATCTGGTGAAGTTTTTCCTCTGTCATTACGTCATGG
TCGATTTCAATTTCTATTGATGCTTTCCAGTCGTAATCAATGATGTATTTTTTGATGTTTGACATCTGTTCATATCC
TCACAGATAAAAAATCGCCCTCACACTGGAGGGCAAAGAAGATTTCCAATAATCAGAACAAGTCGGCTCCTGTTTAG
TTACGAGCGACATTGCTCCGTGTATTCACTCGTTGGAATGAATACACAGTGCAGTGTTTATTCTGTTATTTATGCCA
AAAATAAAGGCCACTATCAGGCAGCTTTGTTGTTCTGTTTACCAAGTTCTCTGGCAATCATTGCCGTCGTTCGTATT
GCCCATTTATCGACATATTTCCCATCTTCCATTACAGGAAACATTTCTTCAGGCTTAACCATGCATTCCGATTGCAG
CTTGCATCCATTGCATCGCTTGAATTGTCCACACCATTGATTTTTATCAATAGTCGTAGTCATACGGATAGTCCTGG
TATTGTTCCATCACATCCTGAGGATGCTCTTCGAACTCTTCAAATTCTTCTTCCATATATCACCTTAAATAGTGGAT
TGCGGTAGTAAAGATTGTGCCTGTCTTTTAACCACATCAGGCTCGGTGGTTCTCGTGTACCCCTACAGCGAGAAATC
GGATAAACTATTACAACCCCTACAGTTTGATGAGTATAGAAATGGATCCACTCGTTATTCTCGGACGAGTGTTCAGT
AATGAACCTCTGGAGAGAACCATGTATATGATCGTTATCTGGGTTGGACTTCTGCTTTTAAGCCCAGATAACTGGCC
TGAATATGTTAATGAGAGAATCGGTATTCCTCATGTGTGGCATGTTTTCGTCTTTGCTCTTGCATTTTCGCTAGCAA
TTAATGTGCATCGATTATCAGCTATTGCCAGCGCCAGATATAAGCGATTTAAGCTAAGAAAACGCATTAAGATGCAA
AACGATAAAGTGCGATCAGTAATTCAAAACCTTACAGAAGAGCAATCTATGGTTTTGTGCGCAGCCCTTAATGAAGG
CAGGAAGTATGTGGTTACATCAAAACAATTCCCATACATTAGTGAGTTGATTGAGCTTGGTGTGTTGAACAAAACTT
TTTCCCGATGGAATGGAAAGCATATATTATTCCCTATTGAGGATATTTACTGGACTGAATTAGTTGCCAGCTATGAT
CCATATAATATTGAGATAAAGCCAAGGCCAATATCTAAGTAACTAGATAAGAGGAATCGATTTTCCCTTAATTTTCT

GGCGTCCACTGCATGTTATGCCGCGTTCGCCAGGCTTGCTGTACCATGTGCGCTGATTCTTGCGCTCAATACGTTGC
AGGTTGCTTTCAATCTGTTTGTGGTATTCAGCCAGCACTGTAAGGTCTATCGGATTTAGTGCGCTTTCTACTCGTGA
TTTCGGTTTGCGATTCAGCGAGAGAATAGGGCGGTTAACTGGTTTTGCGCTTACCCCAACCAACAGGGGATTTGCTG
CTTTCCATTGAGCCTGTTTCTCTGCGCGACGTTCGCGGCGGCGTGTTTGTGCATCCATCTGGATTCTCCTGTCAGTT
AGCTTTGGTGGTGTGTGGCAGTTGTAGTCCTGAACGAAAACCCCCCGCGATTGGCACATTGGCAGCTAATCCGGAAT
CGCACTTACGGCCAATGCTTCGTTTCGTATCACACACCCCAAAGCCTTCTGCTTTGAATGCTGCCCTTCTTCAGGGC
TTAATTTTTAAGAGCGTCACCTTCATGGTGGTCAGTGCGTCCTGCTGATGTGCTCAGTATCACCGCCAGTGGTATTT
ATGTCAACACCGCCAGAGATAATTTATCACCGCAGATGGTTATCTGTATGTTTTTTATATGAATTTATTTTTTGCAG
GGGGGCATTGTTTGGTAGGTGAGAGATCTGAATTGCTATGTTTAGTGAGTTGTATCTATTTATTTTTCAATAAATAC
AATTGGTTATGTGTTTTGGGGGCGATCGTGAGGCAAAGAAAACCCGGCGCTGAGGCCGGGTTATTCTTGTTCTCTGG
TCAAATTATATAGTTGGAAAACAAGGATGCATATATGAATGAACGATGCAGAGGCAATGCCGATGGCGATAGTGGGT
ATCATGTAGCCGCTTATGCTGGAAAGAAGCAATAACCCGCAGAAAAACAAAGCTCCAAGCTCAACAAAACTAAGGGC
ATAGACAATAACTACCGATGTCATATACCCATACTCTCTAATCTTGGCCAGTCGGCGCGTTCTGCTTCCGATTAGAA
ACGTCAAGGCAGCAATCAGGATTGCAATCATGGTTCCTGCATATGATGACAATGTCGCCCCAAGACCATCTCTATGA
GCTGAAAAAGAAACACCAGGAATGTAGTGGCGGAAAAGGAGATAGCAAATGCTTACGATAACGTAAGGAATTATTAC
TATGTAAACACCAGGCATGATTCTGTTCCGCATAATTACTCCTGATAATTAATCCTTAACTTTGCCCACCTGCCTTT
TAAAACATTCCAGTATATCACTTTTCATTCTTGCGTAGCAATATGCCATCTCTTCAGCTATCTCAGCATTGGTGACC
TTGTTCAGAGGCGCTGAGAGATGGCCTTTTTCTGATAGATAATGTTCTGTTAAAATATCTCCGGCCTCATCTTTTGC
CCGCAGGCTAATGTCTGAAAATTGAGGTGACGGGTTAAAAATAATATCCTTGGCAACCTTTTTTATATCCCTTTTAA
ATTTTGGCTTAATGACTATATCCAATGAGTCAAAAAGCTCCCCTTCAATATCTGTTGCCCCTAAGACCTTTAATATA
TCGCCAAATACAGGTAGCTTGGCTTCTACCTTCACCGTTGTTCGGCCGATGAAATGCATATGCATAACATCGTCTTT
GGTGGTTCCCCTCATCAGTGGCTCTATCTGAACGCGCTCTCCACTGCTTAATGACATTCCTTTCCCGATTAAAAAAT
CTGTCAGATCGGATGTGGTCGGCCCGAAAACAGTTCTGGCAAAACCAATGGTGTCGCCTTCAACAAACAAAAAAGAT
GGGAATCCCAATGATTCGTCATCTGCGAGGCTGTTCTTAATATCTTCAACTGAAGCTTTAGAGCGATTTATCTTCTG
AACCAGACTCTTGTCATTTGTTTTGGTAAAGAGAAAAGTTTTTCCATCGATTTTATGAATATACAAATAATTGGAGC
CAACCTGCAGGTGATGATTATCAGCCAGCAGAGAATTAAGGAAAACAGACAGGTTTATTGAGCGCTTATCTTTCCCT
TTATTTTTGCTGCGGTAAGTCGCATAAAAACCATTCTTCATAATTCAATCCATTTACTATGTTATGTTCTGAGGGGA
GTGAAAATTCCCCTAATTCGATGAAGATTCTTGCTCAATTGTTATCAGCTATGCGCCGACCAGAACACCTTGCCGAT
CAGCCAAACGTCTCTTCAGGCCACTGACTAGCGATAACTTTCCCCACAACGGAACAACTCTCATTGCATGGGATCAT
TGGGTACTGTGGGTTTAGTGGTTGTAAAAACACCTGACCGCTATCCCTGATCAGTTTCTTGAAGGTAAACTCATCAC
CCCCAAGTCTGGCTATGCAGAAATCACCTGGCTCAACAGCCTGCTCAGGGTCAACGAGAATTAACATTCCGTCAGGA
AAGCTTGGCTTGGAGCCTGTTGGTGCGGTCATGGAATTACCTTCAACCTCAAGCCAGAATGCAGAATCACTGGCTTT
TTTGGTTGTGCTTACCCATCTCTCCGCATCACCTTTGGTAAAGGTTCTAAGCTTAGGTGAGAACATCCCTGCCTGAA
CATGAGAAAAAACAGGGTACTCATACTCACTTCTAAGTGACGGCTGCATACTAACCGCTTCATACATCTCGTAGATT
TCTCTGGCGATTGAAGGGCTAAATTCTTCAACGCTAACTTTGAGAATTTTTGTAAGCAATGCGGCGTTATAAGCATT
TAATGCATTGATGCCATTAAATAAAGCACCAACGCCTGACTGCCCCATCCCCATCTTGTCTGCGACAGATTCCTGGG
ATAAGCCAAGTTCATTTTTCTTTTTTTCATAAATTGCTTTAAGGCGACGTGCGTCCTCAAGCTGCTCTTGTGTTAAT
GGTTTCTTTTTTGTGCTCATACGTTAAATCTATCACCGCAAGGGATAAATATCTAACACCGTGCGTGTTGACTATTT
TACCTCTGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATGGAACAACGCATAACCCTGAAAGATTATGCAA
TGCGCTTTGGGCAAACCAAGACAGCTAAAGATCTCGGCGTATATCAAAGCGCGATCAACAAGGCCATTCATGCAGGC
CGAAAGATTTTTTTAACTATAAACGCTGATGGAAGCGTTTATGCGGAAGAGGTAAAGCCCTTCCCGAGTAACAAAAA
AACAACAGCATAAATAACCCCGCTCTTACACATTCCAGCCCTGAAAAAGGGCATCAAATTAAACCACACCTATGGTG
TATGCATTTATTTGCATACATTCAATCAATTGTTATCTAAGGAAATACTTACATATGGTTCGTGCAAACAAACGCAA
CGAGGCTCTACGAATCGAGAGTGCGTTGCTTAACAAAATCGCAATGCTTGGAACTGAGAAGACAGCGGAAGCTGTGG
GCGTTGATAAGTCGCAGATCAGCAGGTGGAAGAGGGACTGGATTCCAAAGTTCTCAATGCTGCTTGCTGTTCTTGAA
TGGGGGGTCGTTGACGACGACATGGCTCGATTGGCGCGACAAGTTGCTGCGATTCTCACCAATAAAAAACGCCCGGC
GGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGGAGTCATTATGAC
AAATACAGCAAAAATACTCAACTTCGGCAGAGGTAACTTTGCCGGACAGGAGCGTAATGTGGCAGATCTCGATGATG
GTTACGCCAGACTATCAAATATGCTGCTTGAGGCTTATTCGGGCGCAGATCTGACCAAGCGACAGTTTAAAGTGCTG
CTTGCCATTCTGCGTAAAACCTATGGGTGGAATAAACCAATGGACAGAATCACCGATTCTCAACTTAGCGAGATTAC
AAAGTTACCTGTCAAACGGTGCAATGAAGCCAAGTTAGAACTCGTCAGAATGAATATTATCAAGCAGCAAGGCGGCA
TGTTTGGACCAAATAAAAACATCTCAGAATGGTGCATCCCTCAAAACGAGGGAAAATCCCCTAAAACGAGGGATAAA
ACATCCCTCAAATTGGGGGATTGCTATCCCTCAAAACAGGGGGACACAAAAGACACTATTACAAAAGAAAAAAGAAA
AGATTATTCGTCAGAGAATTCTGGCGAATCCTCTGACCAGCCAGAAAACGACCTTTCTGTGGTGAAACCGGATGCTG
CAATTCAGAGCGGCAGCAAGTGGGGGACAGCAGAAGACCTGACCGCCGCAGAGTGGATGTTTGACATGGTGAAGACT
ATCGCACCATCAGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCCGCCTGATGCGTGAACGTGACGGACG
TAACCACCGCGACATGTGTGTGCTGTTCCGCTGGGCATGCCAGGACAACTTCTGGTCCGGTAACGTGCTGAGCCCGG

CCAAACTCCGCGATAAGTGGACCCAACTCGAAATCAACCGTAACAAGCAACAGGCAGGCGTGACAGCCAGCAAACCA
AAACTCGACCTGACAAACACAGACTGGATTTACGGGGTGGATCTATGAAAAACATCGCCGCACAGATGGTTAACTTT
GACCGTGAGCAGATGCGTCGGATCGCCAACAACATGCCGGAACAGTACGACGAAAAGCCGCAGGTACAGCAGGTAGC
GCAGATCATCAACGGTGTGTTCAGCCAGTTACTGGCAACTTTCCCGGCGAGCCTGGCTAACCGTGACCAGAACGAAG
TGAACGAAATCCGTCGCCAGTGGGTTCTGGCTTTTCGGGAAAACGGGATCACCACGATGGAACAGGTTAACGCAGGA
ATGCGCGTAGCCCGTCGGCAGAATCGACCATTTCTGCCATCACCCGGGCAGTTTGTTGCATGGTGCCGGGAAGAAGC
ATCCGTTACCGCCGGACTGCCAAACGTCAGCGAGCTGGTTGATATGGTTTACGAGTATTGCCGGAAGCGAGGCCTGT
ATCCGGATGCGGAGTCTTATCCGTGGAAATCAAACGCGCACTACTGGCTGGTTACCAACCTGTATCAGAACATGCGG
GCCAATGCGCTTACTGATGCGGAATTACGCCGTAAGGCCGCAGATGAGCTTGTCCATATGACTGCGAGAATTAACCG
TGGTGAGGCGATCCCTGAACCAGTAAAACAACTTCCTGTCATGGGCGGTAGACCTCTAAATCGTGCACAGGCTCTGG
CGAAGATCGCAGAAATCAAAGCTAAGTTCGGACTGAAAGGAGCAAGTGTATGACGGGCAAAGAGGCAATTATTCATT
ACCTGGGGACGCATAATAGCTTCTGTGCGCCGGACGTTGCCGCGCTAACAGGCGCAACAGTAACCAGCATAAATCAG
GCCGCGGCTAAAATGGCACGGGCAGGTCTTCTGGTTATCGAAGGTAAGGTCTGGCGAACGGTGTATTACCGGTTTGC
TACCAGGGAAGAACGGGAAGGAAAGATGAGCACGAACCTGGTTTTTAAGGAGTGTCGCCAGAGTGCCGCGATGAAAC
GGGTATTGGCGGTATATGGAGTTAAAAGATGACCATCTACATTACTGAGCTAATAACAGGCCTGCTGGTAATCGCAG
GCCTTTTTATTTGGGGGAGAGGGAAGTCATGAAAAAACTAACCTTTGAAATTCGATCTCCAGCACATCAGCAAAACG
CTATTCACGCAGTACAGCAAATCCTTCCAGACCCAACCAAACCAATCGTAGTAACCATTCAGGAACGCAACCGCAGC
TTAGACCAAAACAGGAAGCTATGGGCCTGCTTAGGTGACGTCTCTCGTCAGGTTGAATGGCATGGTCGCTGGCTGGA
TGCAGAAAGCTGGAAGTGTGTGTTTACCGCAGCATTAAAGCAGCAGGATGTTGTTCCTAACCTTGCCGGGAATGGCT
TTGTGGTAATAGGCCAGTCAACCAGCAGGATGCGTGTAGGCGAATTTGCGGAGCTATTAGAGCTTATACAGGCATTC
GGTACAGAGCGTGGCGTTAAGTGGTCAGACGAAGCGAGACTGGCTCTGGAGTGGAAAGCGAGATGGGGAGACAGGGC
TGCATGATAAATGTCGTTAGTTTCTCCGGTGGCAGGACGTCAGCATATTTGCTCTGGCTAATGGAGCAAAAGCGACG
GGCAGGTAAAGACGTGCATTACGTTTTCATGGATACAGGTTGTGAACATCCAATGACATATCGGTTTGTCAGGGAAG
TTGTGAAGTTCTGGGATATACCGCTCACCGTATTGCAGGTTGATATCAACCCGGAGCTTGGACAGCCAAATGGTTAT
ACGGTATGGGAACCAAAGGATATTCAGACGCGAATGCCTGTTCTGAAGCCATTTATCGATATGGTAAAGAAATATGG
CACTCCATACGTCGGCGGCGCGTTCTGCACTGACAGATTAAAACTCGTTCCCTTCACCAAATACTGTGATGACCATT
TCGGGCGAGGGAATTACACCACGTGGATTGGCATCAGAGCTGATGAACCGAAGCGGCTAAAGCCAAAGCCTGGAATC
AGATATCTTGCTGAACTGTCAGACTTTGAGAAGGAAGATATCCTCGCATGGTGGAAGCAACAACCATTCGATTTGCA
AATACCGGAACATCTCGGTAACTGCATATTCTGCATTAAAAAATCAACGCAAAAAATCGGACTTGCCTGCAAAGATG
AGGAGGGATTGCAGCGTGTTTTTAATGAGGTCATCACGGGATCCCATGTGCGTGACGGACATCGGGAAACGCCAAAG
GAGATTATGTACCGAGGAAGAATGTCGCTGGACGGTATCGCGAAAATGTATTCAGAAAATGATTATCAAGCCCTGTA
TCAGGACATGGTACGAGCTAAAAGATTCGATACCGGCTCTTGTTCTGAGTCATGCGAAATATTTGGAGGGCAGCTTG
ATTTCGACTTCGGGAGGGAAGCTGCATGATGCGATGTTATCGGTGCGGTGAATGCAAAGAAGATAACCGCTTCCGAC
CAAATCAACCTTACTGGAATCGATGGTGTCTCCGGTGTGAAAGAACACCAACAGGGGTGTTACCACTACCGCAGGAA
AAGGAGGACGTGTGGCGAGACAGCGACGAAGTATCACCGACATAATCTGCGAAAACTGCAAATACCTTCCAACGAAA
CGCACCAGAAATAAACCCAAGCCAATCCCAAAAGAATCTGACGTAAAAACCTTCAACTACACGGCTCACCTGTGGGA
TATCCGGTGGCTAAGACGTCGTGCGAGGAAAACAAGGTGATTGACCAAAATCGAAGTTACGAACAAGAAAGCGTCGA
GCGAGCTTTAACGTGCGCTAACTGCGGTCAGAAGCTGCATGTGCTGGAAGTTCACGTGTGTGAGCACTGCTGCGCAG
AACTGATGAGCGATCCGAATAGCTCGATGCACGAGGAAGAAGATGATGGCTAAACCAGCGCGAAGACGATGTAAAAA
CGATGAATGCCGGGAATGGTTTCACCCTGCATTCGCTAATCAGTGGTGGTGCTCTCCAGAGTGTGGAACCAAGATAG
CACTCGAACGACGAAGTAAAGAACGCGAAAAAGCGGAAAAAGCAGCAGAGAAGAAACGACGACGAGAGGAGCAGAAA
CAGAAAGATAAACTTAAGATTCGAAAACTCGCCTTAAAGCCCCGCAGTTACTGGATTAAACAAGCCCAACAAGCCGT
AAACGCCTTCATCAGAGAAAGAGACCGCGACTTACCATGTATCTCGTGCGGAACGCTCACGTCTGCTCAGTGGGATG
CCGGACATTACCGGACAACTGCTGCGGCACCTCAACTCCGATTTAATGAACGCAATATTCACAAGCAATGCGTGGTG
TGCAACCAGCACAAAAGCGGAAATCTCGTTCCGTATCGCGTCGAACTGATTAGCCGCATCGGGCAGGAAGCAGTAGA
CGAAATCGAATCAAACCATAACCGCCATCGCTGGACTATCGAAGAGTGCAAGGCGATCAAGGCAGAGTACCAACAGA
AACTCAAAGACCTGCGAAATAGCAGAAGTGAGGCCGCATGACGTTCTCAGTAAAAACCATTCCAGACATGCTCGTTG
AAGCATACGGAAATCAGACAGAAGTAGCACGCAGACTGAAATGTAGTCGCGGTACGGTCAGAAAATACGTTGATGAT
AAAGACGGGAAAATGCACGCCATCGTCAACGACGTTCTCATGGTTCATCGCGGATGGAGTGAAAGAGATGCGCTATT
ACGAAAAAATTGATGGCAGCAAATACCGAAATATTTGGGTAGTTGGCGATCTGCACGGATGCTACACGAACCTGATG
AACAAACTGGATACGATTGGATTCGACAACAAAAAAGACCTGCTTATCTCGGTGGGCGATTTGGTTGATCGTGGTGC
AGAGAACGTTGAATGCCTGGAATTAATCACATTCCCCTGGTTCAGAGCTGTACGTGGAAACCATGAGCAAATGATGA
TTGATGGCTTATCAGAGCGTGGAAACGTTAATCACTGGCTGCTTAATGGCGGTGGCTGGTTCTTTAATCTCGATTAC
GACAAAGAAATTCTGGCTAAAGCTCTTGCCCATAAAGCAGATGAACTTCCGTTAATCATCGAACTGGTGAGCAAAGA
TAAAAAATATGTTATCTGCCACGCCGATTATCCCTTTGACGAATACGAGTTTGGAAAGCCAGTTGATCATCAGCAGG
TAATCTGGAACCGCGAACGAATCAGCAACTCACAAAACGGGATCGTGAAAGAAATCAAAGGCGCGGACACGTTCATC
TTTGGTCATACGCCAGCAGTGAAACCACTCAAGTTTGCCAACCAAATGTATATCGATACCGGCGCAGTGTTCTGCGG

AAACCTAACATTGATTCAGGTACAGGGAGAAGGCGCATGAGACTCGAAAGCGTAGCTAAATTTCATTCGCCAAAAAG
CCCGATGATGAGCGACTCACCACGGGCCACGGCTTCTGACTCTCTTTCCGGTACTGATGTGATGGCTGCTATGGGGA
TGGCGCAATCACAAGCCGGATTCGGTATGGCTGCATTCTGCGGTAAGCACGAACTCAGCCAGAACGACAAACAAAAG
GCTATCAACTATCTGATGCAATTTGCACACAAGGTATCGGGGAAATACCGTGGTGTGGCAAAGCTTGAAGGAAATAC
TAAGGCAAAGGTACTGCAAGTGCTCGCAACATTCGCTTATGCGGATTATTGCCGTAGTGCCGCGACGCCGGGGGCAA
GATGCAGAGATTGCCATGGTACAGGCCGTGCGGTTGATATTGCCAAAACAGAGCTGTGGGGGAGAGTTGTCGAGAAA
GAGTGCGGAAGATGCAAAGGCGTCGGCTATTCAAGGATGCCAGCAAGCGCAGCATATCGCGCTGTGACGATGCTAAT
CCCAAACCTTACCCAACCCACCTGGTCACGCACTGTTAAGCCGCTGTATGACGCTCTGGTGGTGCAATGCCACAAAG
AAGAGTCAATCGCAGACAACATTTTGAATGCGGTCACACGTTAGCAGCATGATTGCCACGGATGGCAACATATTAAC
GGCATGATATTGACTTATTGAATAAAATTGGGTAAATTTGACTCAACGATGGGTTAATTCGCTCGTTGTGGTAGTGA
GATGAAAAGAGGCGGCGCTTACTACCGATTCCGCCTAGTTGGTCACTTCGACGTATCGTCTGGAACTCCAACCATCG
CAGGCAGAGAGGTCTGCAAAATGCAATCCCGAAACAGTTCGCAGGTAATAGTTAGAGCCTGCATAACGGTTTCGGGA
TTTTTTATATCTGCACAACAGGTAAGAGCATTGAGTCGATAATCGTGAAGAGTCGGCGAGCCTGGTTAGCCAGTGCT
CTTTCCGTTGTGCTGAATTAAGCGAATACCGGAAGCAGAACCGGATCACCAAATGCGTACAGGCGTCATCGCCGCCC
AGCAACAGCACAACCCAAACTGAGCCGTAGCCACTGTCTGTCCTGAATTCATTAGTAATAGTTACGCTGCGGCCTTT
TACACATGACCTTCGTGAAAGCGGGTGGCAGGAGGTCGCGCTAACAACCTCCTGCCGTTTTGCCCGTGCATATCGGT
CACGAACAAATCTGATTACTAAACACAGTAGCCTGGATTTGTTCTATCAGTAATCGACCTTATTCCTAATTAAATAG
AGCAAATCCCCTTATTGGGGGTAAGACATGAAGATGCCAGAAAAACATGACCTGTTGGCCGCCATTCTCGCGGCAAA
GGAACAAGGCATCGGGGCAATCCTTGCGTTTGCAATGGCGTACCTTCGCGGCAGATATAATGGCGGTGCGTTTACAA
AAACAGTAATCGACGCAACGATGTGCGCCATTATCGCCTAGTTCATTCGTGACCTTCTCGACTTCGCCGGACTAAGT
AGCAATCTCGCTTATATAACGAGCGTGTTTATCGGCTACATCGGTACTGACTCGATTGGTTCGCTTATCAAACGCTT
CGCTGCTAAAAAAGCCGGAGTAGAAGATGGTAGAAATCAATAATCAACGTAAGGCGTTCCTCGATATGCTGGCGTGG
TCGGAGGGAACTGATAACGGACGTCAGAAAACCAGAAATCATGGTTATGACGTCATTGTAGGCGGAGAGCTATTTAC
TGATTACTCCGATCACCCTCGCAAACTTGTCACGCTAAACCCAAAACTCAAATCAACAGGCGCCGGACGCTACCAGC
TTCTTTCCCGTTGGTGGGATGCCTACCGCAAGCAGCTTGGCCTGAAAGACTTCTCTCCGAAAAGTCAGGACGCTGTG
GCATTGCAGCAGATTAAGGAGCGTGGCGCTTTACCTATGATTGATCGTGGTGATATCCGTCAGGCAATCGACCGTTG
CAGCAATATCTGGGCTTCACTGCCGGGCGCTGGTTATGGTCAGTTCGAGCATAAGGCTGACAGCCTGATTGCAAAAT
TCAAAGAAGCGGGCGGAACGGTCAGAGAGATTGATGTATGAGCAGAGTCACCGCGATTATCTCCGCTCTGGTTATCT
GCATCATCGTCTGCCTGTCATGGGCTGTTAATCATTACCGTGATAACGCCATTACCTACAAAGCCCAGCGCGACAAA
AATGCCAGAGAACTGAAGCTGGCGAACGCGGCAATTACTGACATGCAGATGCGTCAGCGTGATGTTGCTGCGCTCGA
TGCAAAATACACGAAGGAGTTAGCTGATGCTAAAGCTGAAAATGATGCTCTGCGTGATGATGTTGCCGCTGGTCGTC
GTCGGTTGCACATCAAAGCAGTCTGTCAGTCAGTGCGTGAAGCCACCACCGCCTCCGGCGTGGATAATGCAGCCTCC
CCCCGACTGGCAGACACCGCTGAACGGGATTATTTCACCCTCAGAGAGAGGCTGATCACTATGCAAAAACAACTGGA
AGGAACCCAGAAGTATATTAATGAGCAGTGCAGATAGAGTTGCCCATATCGATGGGCAACTCATGCAATTATTGTGA
GCAATACACACGCGCTTCCAGCGGAGTATAAATGCCTAAAGTAATAAAACCGAGCAATCCATTTACGAATGTTTGCT
GGGTTTCTGTTTTAACAACATTTTCTGCGCCGCCACAAATTTTGGCTGCATCGACAGTTTTCTTCTGCCCAATTCCA
GAAACGAAGAAATGATGGGTGATGGTTTCCTTTGGTGCTACTGCTGCCGGTTTGTTTTGAACAGTAAACGTCTGTTG
AGCACATCCTGTAATAAGCAGGGCCAGCGCAGTAGCGAGTAGCATTTTTTTCATGGTGTTATTCCCGATGCTTTTTG
AAGTTCGCAGAATCGTATGTGTAGAAAATTAAACAAACCCTAAACAATGAGTTGAAATTTCATATTGTTAATATTTA
TTAATGTATGTCAGGTGCGATGAATCGTCATTGTATTCCCGGATTAACTATGTCCACAGCCCTGACGGGGAACTTCT
CTGCGGGAGTGTCCGGGAATAATTAAAACGATGCACACAGGGTTTAGCGCGTACACGTATTGCATTATGCCAACGCC
CCGGTGCTGACACGGAAGAAACCGGACGTTATGATTTAGCGTGGAAAGATTTGTGTAGTGTTCTGAATGCTCTCAGT
AAATAGTAATGAATTATCAAAGGTATAGTAATATCTTTTATGTTCATGGATATTTGTAACCCATCGGAAAACTCCTG
CTTTAGCAAGATTTTCCCTGTATTGCTGAAATGTGATTTCTCTTGATTTCAACCTATCATAGGACGTTTCTATAAGA
TGCGTGTTTCTTGAGAATTTAACATTTACAACCTTTTTAAGTCCTTTTATTAACACGGTGTTATCGTTTTCTAACAC
GATGTGAATATTATCTGTGGCTAGATAGTAAATATAATGTGAGACGTTGTGACGTTTTAGTTCAGAATAAAACAATT
CACAGTCTAAATCTTTTCGCACTTGATCGAATATTTCTTTAAAAATGGCAACCTGAGCCATTGGTAAAACCTTCCAT
GTGATACGAGGGCGCGTAGTTTGCATTATCGTTTTTATCGTTTCAATCTGGTCTGACCTCCTTGTGTTTTGTTGATG
ATTTATGTCAAATATTAGGAATGTTTTCACTTAATAGTATTGGTTGCGTAACAAAGTGCGGTCCTGCTGGCATTCTG
GAGGGAAATACAACCGACAGATGTATGTAAGGCCAACGTGCTCAAATCTTCATACAGAAAGATTTGAAGTAATATTT
TAACCGCTAGATGAAGAGCAAGCGCATGGAGCGACAAAATGAATAAAGAACAATCTGCTGATGATCCCTCCGTGGAT
CTGATTCGTGTAAAAAATATGCTTAATAGCACCATTTCTATGAGTTACCCTGATGTTGTAATTGCATGTATAGAACA
TAAGGTGTCTCTGGAAGCATTCAGAGCAATTGAGGCAGCGTTGGTGAAGCACGATAATAATATGAAGGATTATTCCC
TGGTGGTTGACTGATCACCATAACTGCTAATCATTCAAACTATTTAGTCTGTGACAGAGCCAACACGCAGTCTGTCA
CTGTCAGGAAAGTGGTAAAACTGCAACTCAATTACTGCAATGCCCTCGTAATTAAGTGAATTTACAATATCGTCCTG
TTCGGAGGGAAGAACGCGGGATGTTCATTCTTCATCACTTTTAATTGATGTATATGCTCTCTTTTCTGACGTTAGTC
TCCGACGGCAGGCTTCAATGACCCAGGCTGAGAAATTCCCGGACCCTTTTTGCTCAAGAGCGATGTTAATTTGTTCA

ATCATTTGGTTAGGAAAGCGGATGTTGCGGGTTGTTGTTCTGCGGGTTCTGTTCTTCGTTGACATGAGGTTGCCCCG
TATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTTACGTTAAGTTGATGCAGATCAATTAATACGATACCTG
CGTCATAATTGATTATTTGACGTGGTTTGATGGCATATGAATATCCTCCTTAGTTCCTATTCCGAAGTTCCTATTCT
CTAGAAAGTATAGGAACTTCGGCGCGCCTACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCC
TGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTC
ACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG
GAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTC
AGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGC
ACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTACGTATGGCAATGAAA
GACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATC
GCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAA
ACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGT
TTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGA
CAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATG
AATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAGGCGCGCCATTTAAATGAAGTTCCTATTCCGAAGTT
CCTATTCTCTAGAAAGTATAGGAACTTCGAAGCAGCTCCAGCCTACACCTCTGTCGTTTCCTTTCTCTGTTTTTGTC
CGTGGAATGAACAATGGAAGTCAACAAAAAGCAGCTGGCTGACATTTTCGGTGCGAGTATCCGTACCATTCAGAACT
GGCAGGAACAGGGAATGCCCGTTCTGCGAGGCGGTGGCAAGGGTAATGAGGTGCTTTATGACTCTGCCGCCGTCATA
AAATGGTATGCCGAAAGGGATGCTGAAATTGAGAACGAAAAGCTGCGCCGGGAGGTTGAAGAACTGCGGCAGGCCAG
CGAGGCAGATCTCCAGCCAGGAACTATTGAGTACGAACGCCATCGACTTACGCGTGCGCAGGCCGACGCACAGGAAC
TGAAGAATGCCAGAGACTCCGCTGAAGTGGTGGAAACCGCATTCTGTACTTTCGTGCTGTCGCGGATCGCAGGTGAA
ATTGCCAGTATTCTCGACGGGCTCCCCCTGTCGGTGCAGCGGCGTTTTCCGGAACTGGAAAACCGACATGTTGATTT
CCTGAAACGGGATATCATCAAAGCCATGAACAAAGCAGCCGCGCTGGATGAACTGATACCGGGGTTGCTGAGTGAAT
ATATCGAACAGTCAGGTTAACAGGCTGCGGCATTTTGTCCGCGCCGGGCTTCGCTCACTGTTCAGGCCGGAGCCACA
GACCGCCGTTGAATGGGCGGATGCTAATTACTATCTCCCGAAAGAATCCGCATACCAGGAAGGGCGCTGGGAAACAC
TGCCCTTTCAGCGGGCCATCATGAATGCGATGGGCAGCGACTACATCCGTGAGGTGAATGTGGTGAAGTCTGCCCGT
GTCGGTTATTCCAAAATGCTGCTGGGTGTTTATGCCTACTTTATAGAGCATAAGCAGCGCAACACCCTTATCTGGTT
GCCGACGGATGGTGATGCCGAGAACTTTATGAAAACCCACGTTGAGCCGACTATTCGTGATATTCCGTCGCTGCTGG
CGCTGGCCCCGTGGTATGGCAAAAAGCACCGGGATAACACGCTCACCATGAAGCGTTTCACTAATGGGCGTGGCTTC
TGGTGCCTGGGCGGTAAAGCGGCAAAAAACTACCGTGAAAAGTCGGTGGATGTGGCGGGTTATGATGAACTTGCTGC
TTTTGATGATGATATTGAACAGGAAGGCTCTCCGACGTTCCTGGGTGACAAGCGTATTGAAGGCTCGGTCTGGCCAA
AGTCCATCCGTGGCTCCACGCCAAAAGTGAGAGGCACCTGTCAGATTGAGCGTGCAGCCAGTGAATCCCCGCATTTT
ATGCGTTTTCATGTTGCCTGCCCGCATTGCGGGGAGGAGCAGTATCTTAAATTTGGCGACAAAGAGACGCCGTTTGG
CCTCAAATGGACGCCGGATGACCCCTCCAGCGTGTTTTATCTCTGCGAGCATAATGCCTGCGTCATCCGCCAGCAGG
AGCTGGACTTTACTGATGCCCGTTATATCTGCGAAAAGACCGGGATCTGGACCCGTGATGGCATTCTCTGGTTTTCG
TCATCCGGTGAAGAGATTGAGCCACCTGACAGTGTGACCTTTCACATCTGGACAGCGTACAGCCCGTTCACCACCTG
GGTGCAGATTGTCAAAGACTGGATGAAAACGAAAGGGGATACGGGAAAACGTAAAACCTTCGTAAACACCACGCTCG
GTGAGACGTGGGAGGCGAAAATTGGCGAACGTCCGGATGCTGAAGTGATGGCAGAGCGGAAAGAGCATTATTCAGCG
CCCGTTCCTGACCGTGTGGCTTACCTGACCGCCGGTATCGACTCCCAGCTGGACCGCTACGAAATGCGCGTATGGGG
ATGGGGGCCGGGTGAGGAAAGCTGGCTGATTGACCGGCAGATTATTATGGGCCGCCACGACGATGAACAGACGCTGC
TGCGTGTGGATGAGGCCATCAATAAAACCTATACCCGCCGGAATGGTGCAGAAATGTCGATATCCCGTATCTGCTGG
GATACTGGCGGGATTGACCCGACCATTGTGTATGAACGCTCGAAAAAACATGGGCTGTTCCGGGTGATCCCCATTAA
AGGGGCATCCGTCTACGGAAAGCCGGTGGCCAGCATGCCACGTAAGCGAAACAAAAACGGGGTTTACCTTACCGAAA
TCGGTACGGATACCGCGAAAGAGCAGATTTATAACCGCTTCACACTGACGCCGGAAGGGGATGAACCGCTTCCCGGT
GCCGTTCACTTCCCGAATAACCCGGATATTTTTGATCTGACCGAAGCGCAGCAGCTGACTGCTGAAGAGCAGGTCGA
AAAATGGGTGGATGGCAGGAAAAAAATACTGTGGGACAGCAAAAAGCGACGCAATGAGGCACTCGACTGCTTCGTTT
ATGCGCTGGCGGCGCTGCGCATCAGTATTTCCCGCTGGCAGCTGGATCTCAGTGCGCTGCTGGCGAGCCTGCAGGAA
GAGGATGGTGCAGCAACCAACAAGAAAACACTGGCAGATTACGCCCGTGCCTTATCCGGAGAGGATGAATGACGCGA
CAGGAAGAACTTGCCGCTGCCCGTGCGGCACTGCATGACCTGATGACAGGTAAACGGGTGGCAACAGTACAGAAAGA
CGGACGAAGGGTGGAGTTTACGGCCACTTCCGTGTCTGACCTGAAAAAATATATTGCAGAGCTGGAAGTGCAGACCG
GCATGACACAGCGACGCAGGGGACCTGCAGGATTTTATGTATGAAAACGCCCACCATTCCCACCCTTCTGGGGCCGG
ACGGCATGACATCGCTGCGCGAATATGCCGGTTATCACGGCGGTGGCAGCGGATTTGGAGGGCAGTTGCGGTCGTGG
AACCCACCGAGTGAAAGTGTGGATGCAGCCCTGTTGCCCAACTTTACCCGTGGCAATGCCCGCGCAGACGATCTGGT
ACGCAATAACGGCTATGCCGCCAACGCCATCCAGCTGCATCAGGATCATATCGTCGGGTCTTTTTTCCGGCTCAGTC
ATCGCCCAAGCTGGCGCTATCTGGGCATCGGGGAGGAAGAAGCCCGTGCCTTTTCCCGCGAGGTTGAAGCGGCATGG
AAAGAGTTTGCCGAGGATGACTGCTGCTGCATTGACGTTGAGCGAAAACGCACGTTTACCATGATGATTCGGGAAGG
TGTGGCCATGCACGCCTTTAACGGTGAACTGTTCGTTCAGGCCACCTGGGATACCAGTTCGTCGCGGCTTTTCCGGA

CACAGTTCCGGATGGTCAGCCCGAAGCGCATCAGCAACCCGAACAATACCGGCGACAGCCGGAACTGCCGTGCCGGT
GTGCAGATTAATGACAGCGGTGCGGCGCTGGGATATTACGTCAGCGAGGACGGGTATCCTGGCTGGATGCCGCAGAA
ATGGACATGGATACCCCGTGAGTTACCCGGCGGGCGCGCCTCGTTCATTCACGTTTTTGAACCCGTGGAGGACGGGC
AGACTCGCGGTGCAAATGTGTTTTACAGCGTGATGGAGCAGATGAAGATGCTCGACACGCTGCAGAACACGCAGCTG
CAGAGCGCCATTGTGAAGGCGATGTATGCCGCCACCATTGAGAGTGAGCTGGATACGCAGTCAGCGATGGATTTTAT
TCTGGGCGCGAACAGTCAGGAGCAGCGGGAAAGGCTGACCGGCTGGATTGGTGAAATTGCCGCGTATTACGCCGCAG
CGCCGGTCCGGCTGGGAGGCGCAAAAGTACCGCACCTGATGCCGGGTGACTCACTGAACCTGCAGACGGCTCAGGAT
ACGGATAACGGCTACTCCGTGTTTGAGCAGTCACTGCTGCGGTATATCGCTGCCGGGCTGGGTGTCTCGTATGAGCA
GCTTTCCCGGAATTACGCCCAGATGAGCTACTCCACGGCACGGGCCAGTGCGAACGAGTCGTGGGCGTACTTTATGG
GGCGGCGAAAATTCGTCGCATCCCGTCAGGCGAGCCAGATGTTTCTGTGCTGGCTGGAAGAGGCCATCGTTCGCCGC
GTGGTGACGTTACCTTCAAAAGCGCGCTTCAGTTTTCAGGAAGCCCGCAGTGCCTGGGGGAACTGCGACTGGATAGG
CTCCGGTCGTATGGCCATCGATGGTCTGAAAGAAGTTCAGGAAGCGGTGATGCTGATAGAAGCCGGACTGAGTACCT
ACGAGAAAGAGTGCGCAAAACGCGGTGACGACTATCAGGAAATTTTTGCCCAGCAGGTCCGTGAAACGATGGAGCGC
CGTGCAGCCGGTCTTAAACCGCCCGCCTGGGCGGCTGCAGCATTTGAATCCGGGCTGCGACAATCAACAGAGGAGGA
GAAGAGTGACAGCAGAGCTGCGTAATCTCCCGCATATTGCCAGCATGGCCTTTAATGAGCCGCTGATGCTTGAACCC
GCCTATGCGCGGGTTTTCTTTTGTGCGCTTGCAGGCCAGCTTGGGATCAGCAGCCTGACGGATGCGGTGTCCGGCGA
CAGCCTGACTGCCCAGGAGGCACTCGCGACGCTGGCATTATCCGGTGATGATGACGGACCACGACAGGCCCGCAGTT
ATCAGGTCATGAACGGCATCGCCGTGCTGCCGGTGTCCGGCACGCTGGTCAGCCGGACGCGGGCGCTGCAGCCGTAC
TCGGGGATGACCGGTTACAACGGCATTATCGCCCGTCTGCAACAGGCTGCCAGCGATCCGATGGTGGACGGCATTCT
GCTCGATATGGACACGCCCGGCGGGATGGTGGCGGGGGCATTTGACTGCGCTGACATCATCGCCCGTGTGCGTGACA
TAAAACCGGTATGGGCGCTTGCCAACGACATGAACTGCAGTGCAGGTCAGTTGCTTGCCAGTGCCGCCTCCCGGCGT
CTGGTCACGCAGACCGCCCGGACAGGCTCCATCGGCGTCATGATGGCTCACAGTAATTACGGTGCTGCGCTGGAGAA
ACAGGGTGTGGAAATCACGCTGATTTACAGCGGCAGCCATAAGGTGGATGGCAACCCCTACAGCCATCTTCCGGATG
ACGTCCGGGAGACACTGCAGTCCCGGATGGACGCAACCCGCCAGATGTTTGCGCAGAAGGTGTCGGCATATACCGGC
CTGTCCGTGCAGGTTGTGCTGGATACCGAGGCTGCAGTGTACAGCGGTCAGGAGGCCATTGATGCCGGACTGGCTGA
TGAACTTGTTAACAGCACCGATGCGATCACCGTCATGCGTGATGCACTGGATGCACGTAAATCCCGTCTCTCAGGAG
GGCGAATGACCAAAGAGACTCAATCAACAACTGTTTCAGCCACTGCTTCGCAGGCTGACGTTACTGACGTGGTGCCA
GCGACGGAGGGCGAGAACGCCAGCGCGGCGCAGCCGGACGTGAACGCGCAGATCACCGCAGCGGTTGCGGCAGAAAA
CAGCCGCATTATGGGGATCCTCAACTGTGAGGAGGCTCACGGACGCGAAGAACAGGCACGCGTGCTGGCAGAAACCC
CCGGTATGACCGTGAAAACGGCCCGCCGCATTCTGGCCGCAGCACCACAGAGTGCACAGGCGCGCAGTGACACTGCG
CTGGATCGTCTGATGCAGGGGGCACCGGCACCGCTGGCTGCAGGTAACCCGGCATCTGATGCCGTTAACGATTTGCT
GAACACACCAGTGTAAGGGATGTTTATGACGAGCAAAGAAACCTTTACCCATTACCAGCCGCAGGGCAACAGTGACC
CGGCTCATACCGCAACCGCGCCCGGCGGATTGAGTGCGAAAGCGCCTGCAATGACCCCGCTGATGCTGGACACCTCC
AGCCGTAAGCTGGTTGCGTGGGATGGCACCACCGACGGTGCTGCCGTTGGCATTCTTGCGGTTGCTGCTGACCAGAC
CAGCACCACGCTGACGTTCTACAAGTCCGGCACGTTCCGTTATGAGGATGTGCTCTGGCCGGAGGCTGCCAGCGACG
AGACGAAAAAACGGACCGCGTTTGCCGGAACGGCAATCAGCATCGTTTAACTTTACCCTTCATCACTAAAGGCCGCC
TGTGCGGCTTTTTTTACGGGATTTTTTTATGTCGATGTACACAACCGCCCAACTGCTGGCGGCAAATGAGCAGAAAT
TTAAGTTTGATCCGCTGTTTCTGCGTCTCTTTTTCCGTGAGAGCTATCCCTTCACCACGGAGAAAGTCTATCTCTCA
CAAATTCCGGGACTGGTAAACATGGCGCTGTACGTTTCGCCGATTGTTTCCGGTGAGGTTATCCGTTCCCGTGGCGG
CTCCACCTCTGAATTTACGCCGGGATATGTCAAGCCGAAGCATGAAGTGAATCCGCAGATGACCCTGCGTCGCCTGC
CGGATGAAGATCCGCAGAATCTGGCGGACCCGGCTTACCGCCGCCGTCGCATCATCATGCAGAACATGCGTGACGAA
GAGCTGGCCATTGCTCAGGTCGAAGAGATGCAGGCAGTTTCTGCCGTGCTTAAGGGCAAATACACCATGACCGGTGA
AGCCTTCGATCCGGTTGAGGTGGATATGGGCCGCAGTGAGGAGAATAACATCACGCAGTCCGGCGGCACGGAGTGGA
GCAAGCGTGACAAGTCCACGTATGACCCGACCGACGATATCGAAGCCTACGCGCTGAACGCCAGCGGTGTGGTGAAT
ATCATCGTGTTCGATCCGAAAGGCTGGGCGCTGTTCCGTTCCTTCAAAGCCGTCAAGGAGAAGCTGGATACCCGTCG
TGGCTCTAATTCCGAGCTGGAGACAGCGGTGAAAGACCTGGGCAAAGCGGTGTCCTATAAGGGGATGTATGGCGATG
TGGCCATCGTCGTGTATTCCGGACAGTACGTGGAAAACGGCGTCAAAAAGAACTTCCTGCCGGACAACACGATGGTG
CTGGGGAACACTCAGGCACGCGGTCTGCGCACCTATGGCTGCATTCAGGATGCGGACGCACAGCGCGAAGGCATTAA
CGCCTCTGCCCGTTACCCGAAAAACTGGGTGACCACCGGCGATCCGGCGCGTGAGTTCACCATGATTCAGTCAGCAC
CGCTGATGCTGCTGGCTGACCCTGATGAGTTCGTGTCCGTACAACTGGCGTAATCATGGCCCTTCGGGGCCATTGTT
TCTCTGTGGAGGAGTCCATGACGAAAGATGAACTGATTGCCCGTCTCCGCTCGCTGGGTGAACAACTGAACCGTGAT
GTCAGCCTGACGGGGACGAAAGAAGAACTGGCGCTCCGTGTGGCAGAGCTGAAAGAGGAGCTTGATGACACGGATGA
AACTGCCGGTCAGGACACCCCTCTCAGCCGGGAAAATGTGCTGACCGGACATGAAAATGAGGTGGGATCAGCGCAGC
CGGATACCGTGATTCTGGATACGTCTGAACTGGTCACGGTCGTGGCACTGGTGAAGCTGCATACTGATGCACTTCAC
GCCACGCGGGATGAACCTGTGGCATTTGTGCTGCCGGGAACGGCGTTTCGTGTCTCTGCCGGTGTGGCAGCCGAAAT
GACAGAGCGCGGCCTGGCCAGAATGCAATAACGGGAGGCGCTGTGGCTGATTTCGATAACCTGTTCGATGCTGCCAT
TGCCCGCGCCGATGAAACGATACGCGGGTACATGGGAACGTCAGCCACCATTACATCCGGTGAGCAGTCAGGTGCGG

TGATACGTGGTGTTTTTGATGACCCTGAAAATATCAGCTATGCCGGACAGGGCGTGCGCGTTGAAGGCTCCAGCCCG
TCCCTGTTTGTCCGGACTGATGAGGTGCGGCAGCTGCGGCGTGGAGACACGCTGACCATCGGTGAGGAAAATTTCTG
GGTAGATCGGGTTTCGCCGGATGATGGCGGAAGTTGTCATCTCTGGCTTGGACGGGGCGTACCGCCTGCCGTTAACC
GTCGCCGCTGAAAGGGGGATGTATGGCCATAAAAGGTCTTGAGCAGGCCGTTGAAAACCTCAGCCGTATCAGCAAAA
CGGCGGTGCCTGGTGCCGCCGCAATGGCCATTAACCGCGTTGCTTCATCCGCGATATCGCAGTCGGCGTCACAGGTT
GCCCGTGAGACAAAGGTACGCCGGAAACTGGTAAAGGAAAGGGCCAGGCTGAAAAGGGCCACGGTCAAAAATCCGCA
GGCCAGAATCAAAGTTAACCGGGGGGATTTGCCCGTAATCAAGCTGGGTAATGCGCGGGTTGTCCTTTCGCGCCGCA
GGCGTCGTAAAAAGGGGCAGCGTTCATCCCTGAAAGGTGGCGGCAGCGTGCTTGTGGTGGGTAACCGTCGTATTCCC
GGCGCGTTTATTCAGCAACTGAAAAATGGCCGGTGGCATGTCATGCAGCGTGTGGCTGGGAAAAACCGTTACCCCAT
TGATGTGGTGAAAATCCCGATGGCGGTGCCGCTGACCACGGCGTTTAAACAAAATATTGAGCGGATACGGCGTGAAC
GTCTTCCGAAAGAGCTGGGCTATGCGCTGCAGCATCAACTGAGGATGGTAATAAAGCGATGAAACATACTGAACTCC
GTGCAGCCGTACTGGATGCACTGGAGAAGCATGACACCGGGGCGACGTTTTTTGATGGTCGCCCCGCTGTTTTTGAT
GAGGCGGATTTTCCGGCAGTTGCCGTTTATCTCACCGGCGCTGAATACACGGGCGAAGAGCTGGACAGCGATACCTG
GCAGGCGGAGCTGCATATCGAAGTTTTCCTGCCTGCTCAGGTGCCGGATTCAGAGCTGGATGCGTGGATGGAGTCCC
GGATTTATCCGGTGATGAGCGATATCCCGGCACTGTCAGATTTGATCACCAGTATGGTGGCCAGCGGCTATGACTAC
CGGCGCGACGATGATGCGGGCTTGTGGAGTTCAGCCGATCTGACTTATGTCATTACCTATGAAATGTGAGGACGCTA
TGCCTGTACCAAATCCTACAATGCCGGTGAAAGGTGCCGGGACCACCCTGTGGGTTTATAAGGGGAGCGGTGACCCT
TACGCGAATCCGCTTTCAGACGTTGACTGGTCGCGTCTGGCAAAAGTTAAAGACCTGACGCCCGGCGAACTGACCGC
TGAGTCCTATGACGACAGCTATCTCGATGATGAAGATGCAGACTGGACTGCGACCGGGCAGGGGCAGAAATCTGCCG
GAGATACCAGCTTCACGCTGGCGTGGATGCCCGGAGAGCAGGGGCAGCAGGCGCTGCTGGCGTGGTTTAATGAAGGC
GATACCCGTGCCTATAAAATCCGCTTCCCGAACGGCACGGTCGATGTGTTCCGTGGCTGGGTCAGCAGTATCGGTAA
GGCGGTGACGGCGAAGGAAGTGATCACCCGCACGGTGAAAGTCACCAATGTGGGACGTCCGTCGATGGCAGAAGATC
GCAGCACGGTAACAGCGGCAACCGGCATGACCGTGACGCCTGCCAGCACCTCGGTGGTGAAAGGGCAGAGCACCACG
CTGACCGTGGCCTTCCAGCCGGAGGGCGTAACCGACAAGAGCTTTCGTGCGGTGTCTGCGGATAAAACAAAAGCCAC
CGTGTCGGTCAGTGGTATGACCATCACCGTGAACGGCGTTGCTGCAGGCAAGGTCAACATTCCGGTTGTATCCGGTA
ATGGTGAGTTTGCTGCGGTTGCAGAAATTACCGTCACCGCCAGTTAATCCGGAGAGTCAGCGATGTTCCTGAAAACC
GAATCATTTGAACATAACGGTGTGACCGTCACGCTTTCTGAACTGTCAGCCCTGCAGCGCATTGAGCATCTCGCCCT
GATGAAACGGCAGGCAGAACAGGCGGAGTCAGACAGCAACCGGAAGTTTACTGTGGAAGACGCCATCAGAACCGGCG
CGTTTCTGGTGGCGATGTCCCTGTGGCATAACCATCCGCAGAAGACGCAGATGCCGTCCATGAATGAAGCCGTTAAA
CAGATTGAGCAGGAAGTGCTTACCACCTGGCCCACGGAGGCAATTTCTCATGCTGAAAACGTGGTGTACCGGCTGTC
TGGTATGTATGAGTTTGTGGTGAATAATGCCCCTGAACAGACAGAGGACGCCGGGCCCGCAGAGCCTGTTTCTGCGG
GAAAGTGTTCGACGGTGAGCTGAGTTTTGCCCTGAAACTGGCGCGTGAGATGGGGCGACCCGACTGGCGTGCCATGC
TTGCCGGGATGTCATCCACGGAGTATGCCGACTGGCACCGCTTTTACAGTACCCATTATTTTCATGATGTTCTGCTG
GATATGCACTTTTCCGGGCTGACGTACACCGTGCTCAGCCTGTTTTTCAGCGATCCGGATATGCATCCGCTGGATTT
CAGTCTGCTGAACCGGCGCGAGGCTGACGAAGAGCCTGAAGATGATGTGCTGATGCAGAAAGCGGCAGGGCTTGCCG
GAGGTGTCCGCTTTGGCCCGGACGGGAATGAAGTTATCCCCGCTTCCCCGGATGTGGCGGACATGACGGAGGATGAC
GTAATGCTGATGACAGTATCAGAAGGGATCGCAGGAGGAGTCCGGTATGGCTGAACCGGTAGGCGATCTGGTCGTTG
ATTTGAGTCTGGATGCGGCCAGATTTGACGAGCAGATGGCCAGAGTCAGGCGTCATTTTTCTGGTACGGAAAGTGAT
GCGAAAAAAACAGCGGCAGTCGTTGAACAGTCGCTGAGCCGACAGGCGCTGGCTGCACAGAAAGCGGGGATTTCCGT
CGGGCAGTATAAAGCCGCCATGCGTATGCTGCCTGCACAGTTCACCGACGTGGCCACGCAGCTTGCAGGCGGGCAAA
GTCCGTGGCTGATCCTGCTGCAACAGGGGGGGCAGGTGAAGGACTCCTTCGGCGGGATGATCCCCATGTTCAGGGGG
CTTGCCGGTGCGATCACCCTGCCGATGGTGGGGGCCACCTCGCTGGCGGTGGCGACCGGTGCGCTGGCGTATGCCTG
GTATCAGGGCAACTCAACCCTGTCCGATTTCAACAAAACGCTGGTCCTTTCCGGCAATCAGGCGGGACTGACGGCAG
ATCGTATGCTGGTCCTGTCCAGAGCCGGGCAGGCGGCAGGGCTGACGTTTAACCAGACCAGCGAGTCACTCAGCGCA
CTGGTTAAGGCGGGGGTAAGCGGTGAGGCTCAGATTGCGTCCATCAGCCAGAGTGTGGCGCGTTTCTCCTCTGCATC
CGGCGTGGAGGTGGACAAGGTCGCTGAAGCCTTCGGGAAGCTGACCACAGACCCGACGTCGGGGCTGACGGCGATGG
CTCGCCAGTTCCATAACGTGTCGGCGGAGCAGATTGCGTATGTTGCTCAGTTGCAGCGTTCCGGCGATGAAGCCGGG
GCATTGCAGGCGGCGAACGAGGCCGCAACGAAAGGGTTTGATGACCAGACCCGCCGCCTGAAAGAGAACATGGGCAC
GCTGGAGACCTGGGCAGACAGGACTGCGCGGGCATTCAAATCCATGTGGGATGCGGTGCTGGATATTGGTCGTCCTG
ATACCGCGCAGGAGATGCTGATTAAGGCAGAGGCTGCGTATAAGAAAGCAGACGACATCTGGAATCTGCGCAAGGAT
GATTATTTTGTTAACGATGAAGCGCGGGCGCGTTACTGGGATGATCGTGAAAAGGCCCGTCTTGCGCTTGAAGCCGC
CCGAAAGAAGGCTGAGCAGCAGACTCAACAGGACAAAAATGCGCAGCAGCAGAGCGATACCGAAGCGTCACGGCTGA
AATATACCGAAGAGGCGCAGAAGGCTTACGAACGGCTGCAGACGCCGCTGGAGAAATATACCGCCCGTCAGGAAGAA
CTGAACAAGGCACTGAAAGACGGGAAAATCCTGCAGGCGGATTACAACACGCTGATGGCGGCGGCGAAAAAGGATTA
TGAAGCGACGCTGAAAAAGCCGAAACAGTCCAGCGTGAAGGTGTCTGCGGGCGATCGTCAGGAAGACAGTGCTCATG
CTGCCCTGCTGACGCTTCAGGCAGAACTCCGGACGCTGGAGAAGCATGCCGGAGCAAATGAGAAAATCAGCCAGCAG
CGCCGGGATTTGTGGAAGGCGGAGAGTCAGTTCGCGGTACTGGAGGAGGCGGCGCAACGTCGCCAGCTGTCTGCACA

GGAGAAATCCCTGCTGGCGCATAAAGATGAGACGCTGGAGTACAAACGCCAGCTGGCTGCACTTGGCGACAAGGTTA
CGTATCAGGAGCGCCTGAACGCGCTGGCGCAGCAGGCGGATAAATTCGCACAGCAGCAACGGGCAAAACGGGCCGCC
ATTGATGCGAAAAGCCGGGGGCTGACTGACCGGCAGGCAGAACGGGAAGCCACGGAACAGCGCCTGAAGGAACAGTA
TGGCGATAATCCGCTGGCGCTGAATAACGTCATGTCAGAGCAGAAAAAGACCTGGGCGGCTGAAGACCAGCTTCGCG
GGAACTGGATGGCAGGCCTGAAGTCCGGCTGGAGTGAGTGGGAAGAGAGCGCCACGGACAGTATGTCGCAGGTAAAA
AGTGCAGCCACGCAGACCTTTGATGGTATTGCACAGAATATGGCGGCGATGCTGACCGGCAGTGAGCAGAACTGGCG
CAGCTTCACCCGTTCCGTGCTGTCCATGATGACAGAAATTCTGCTTAAGCAGGCAATGGTGGGGATTGTCGGGAGTA
TCGGCAGCGCCATTGGCGGGGCTGTTGGTGGCGGCGCATCCGCGTCAGGCGGTACAGCCATTCAGGCCGCTGCGGCG
AAATTCCATTTTGCAACCGGAGGATTTACGGGAACCGGCGGCAAATATGAGCCAGCGGGGATTGTTCACCGTGGTGA
GTTTGTCTTCACGAAGGAGGCAACCAGCCGGATTGGCGTGGGGAATCTTTACCGGCTGATGCGCGGCTATGCCACCG
GCGGTTATGTCGGTACACCGGGCAGCATGGCAGACAGCCGGTCGCAGGCGTCCGGGACGTTTGAGCAGAATAACCAT
GTGGTGATTAACAACGACGGCACGAACGGGCAGATAGGTCCGGCTGCTCTGAAGGCGGTGTATGACATGGCCCGCAA
GGGTGCCCGTGATGAAATTCAGACACAGATGCGTGATGGTGGCCTGTTCTCCGGAGGTGGACGATGAAGACCTTCCG
CTGGAAAGTGAAACCCGGTATGGATGTGGCTTCGGTCCCTTCTGTAAGAAAGGTGCGCTTTGGTGATGGCTATTCTC
AGCGAGCGCCTGCCGGGCTGAATGCCAACCTGAAAACGTACAGCGTGACGCTTTCTGTCCCCCGTGAGGAGGCCACG
GTACTGGAGTCGTTTCTGGAAGAGCACGGGGGCTGGAAATCCTTTCTGTGGACGCCGCCTTATGAGTGGCGGCAGAT
AAAGGTGACCTGCGCAAAATGGTCGTCGCGGGTCAGTATGCTGCGTGTTGAGTTCAGCGCAGAGTTTGAACAGGTGG
TGAACTGATGCAGGATATCCGGCAGGAAACACTGAATGAATGCACCCGTGCGGAGCAGTCGGCCAGCGTGGTGCTCT
GGGAAATCGACCTGACAGAGGTCGGTGGAGAACGTTATTTTTTCTGTAATGAGCAGAACGAAAAAGGTGAGCCGGTC
ACCTGGCAGGGGCGACAGTATCAGCCGTATCCCATTCAGGGGAGCGGTTTTGAACTGAATGGCAAAGGCACCAGTAC
GCGCCCCACGCTGACGGTTTCTAACCTGTACGGTATGGTCACCGGGATGGCGGAAGATATGCAGAGTCTGGTCGGCG
GAACGGTGGTCCGGCGTAAGGTTTACGCCCGTTTTCTGGATGCGGTGAACTTCGTCAACGGAAACAGTTACGCCGAT
CCGGAGCAGGAGGTGATCAGCCGCTGGCGCATTGAGCAGTGCAGCGAACTGAGCGCGGTGAGTGCCTCCTTTGTACT
GTCCACGCCGACGGAAACGGATGGCGCTGTTTTTCCGGGACGTATCATGCTGGCCAACACCTGCACCTGGACCTATC
GCGGTGACGAGTGCGGTTATAGCGGTCCGGCTGTCGCGGATGAATATGACCAGCCAACGTCCGATATCACGAAGGAT
AAATGCAGCAAATGCCTGAGCGGTTGTAAGTTCCGCAATAACGTCGGCAACTTTGGCGGCTTCCTTTCCATTAACAA
ACTTTCGCAGTAAATCCCATGACACAGACAGAATCAGCGATTCTGGCGCACGCCCGGCGATGTGCGCCAGCGGAGTC
GTGCGGCTTCGTGGTAAGCACGCCGGAGGGGGAAAGATATTTCCCCTGCGTGAATATCTCCGGTGAGCCGGAGGCTA
TTTCCGTATGTCGCCGGAAGACTGGCTGCAGGCAGAAATGCAGGGTGAGATTGTGGCGCTGGTCCACAGCCACCCCG
GTGGTCTGCCCTGGCTGAGTGAGGCCGACCGGCGGCTGCAGGTGCAGAGTGATTTGCCGTGGTGGCTGGTCTGCCGG
GGGACGATTCATAAGTTCCGCTGTGTGCCGCATCTCACCGGGCGGCGCTTTGAGCACGGTGTGACGGACTGTTACAC
ACTGTTCCGGGATGCTTATCATCTGGCGGGGATTGAGATGCCGGACTTTCATCGTGAGGATGACTGGTGGCGTAACG
GCCAGAATCTCTATCTGGATAATCTGGAGGCGACGGGGCTGTATCAGGTGCCGTTGTCAGCGGCACAGCCGGGCGAT
GTGCTGCTGTGCTGTTTTGGTTCATCAGTGCCGAATCACGCCGCAATTTACTGCGGCGACGGCGAGCTGCTGCACCA
TATTCCTGAACAACTGAGCAAACGAGAGAGGTACACCGACAAATGGCAGCGACGCACACACTCCCTCTGGCGTCACC
GGGCATGGCGCGCATCTGCCTTTACGGGGATTTACAACGATTTGGTCGCCGCATCGACCTTCGTGTGAAAACGGGGG
CTGAAGCCATCCGGGCACTGGCCACACAGCTCCCGGCGTTTCGTCAGAAACTGAGCGACGGCTGGTATCAGGTACGG
ATTGCCGGGCGGGACGTCAGCACGTCCGGGTTAACGGCGCAGTTACATGAGACTCTGCCTGATGGCGCTGTAATTCA
TATTGTTCCCAGAGTCGCCGGGGCCAAGTCAGGTGGCGTATTCCAGATTGTCCTGGGGGCTGCCGCCATTGCCGGAT
CATTCTTTACCGCCGGAGCCACCCTTGCAGCATGGGGGGCAGCCATTGGGGCCGGTGGTATGACCGGCATCCTGTTT
TCTCTCGGTGCCAGTATGGTGCTCGGTGGTGTGGCGCAGATGCTGGCACCGAAAGCCAGAACTCCCCGTATACAGAC
AACGGATAACGGTAAGCAGAACACCTATTTCTCCTCACTGGATAACATGGTTGCCCAGGGCAATGTTCTGCCTGTTC
TGTACGGGGAAATGCGCGTGGGGTCACGCGTGGTTTCTCAGGAGATCAGCACGGCAGACGAAGGGGACGGTGGTCAG
GTTGTGGTGATTGGTCGCTGATGCAAAATGTTTTATGTGAAACCGCCTGCGGGCGGTTTTGTCATTTATGGAGCGTG
AGGAATGGGTAAAGGAAGCAGTAAGGGGCATACCCCGCGCGAAGCGAAGGACAACCTGAAGTCCACGCAGTTGCTGA
GTGTGATCGATGCCATCAGCGAAGGGCCGATTGAAGGTCCGGTGGATGGCTTAAAAAGCGTGCTGCTGAACAGTACG
CCGGTGCTGGACACTGAGGGGAATACCAACATATCCGGTGTCACGGTGGTGTTCCGGGCTGGTGAGCAGGAGCAGAC
TCCGCCGGAGGGATTTGAATCCTCCGGCTCCGAGACGGTGCTGGGTACGGAAGTGAAATATGACACGCCGATCACCC
GCACCATTACGTCTGCAAACATCGACCGTCTGCGCTTTACCTTCGGTGTACAGGCACTGGTGGAAACCACCTCAAAG
GGTGACAGGAATCCGTCGGAAGTCCGCCTGCTGGTTCAGATACAACGTAACGGTGGCTGGGTGACGGAAAAAGACAT
CACCATTAAGGGCAAAACCACCTCGCAGTATCTGGCCTCGGTGGTGATGGGTAACCTGCCGCCGCGCCCGTTTAATA
TCCGGATGCGCAGGATGACGCCGGACAGCACCACAGACCAGCTGCAGAACAAAACGCTCTGGTCGTCATACACTGAA
ATCATCGATGTGAAACAGTGCTACCCGAACACGGCACTGGTCGGCGTGCAGGTGGACTCGGAGCAGTTCGGCAGCCA
GCAGGTGAGCCGTAATTATCATCTGCGCGGGCGTATTCTGCAGGTGCCGTCGAACTATAACCCGCAGACGCGGCAAT
ACAGCGGTATCTGGGACGGAACGTTTAAACCGGCATACAGCAACAACATGGCCTGGTGTCTGTGGGATATGCTGACC
CATCCGCGCTACGGCATGGGGAAACGTCTTGGTGCGGCGGATGTGGATAAATGGGCGCTGTATGTCATCGGCCAGTA
CTGCGACCAGTCAGTGCCGGACGGCTTTGGCGGCACGGAGCCGCGCATCACCTGTAATGCGTACCTGACCACACAGC

GTAAGGCGTGGGATGTGCTCAGCGATTTCTGCTCGGCGATGCGCTGTATGCCGGTATGGAACGGGCAGACGCTGACG
TTCGTGCAGGACCGACCGTCGGATAAGACGTGGACCTATAACCGCAGTAATGTGGTGATGCCGGATGATGGCGCGCC
GTTCCGCTACAGCTTCAGCGCCCTGAAGGACCGCCATAATGCCGTTGAGGTGAACTGGATTGACCCGAACAACGGCT
GGGAGACGGCGACAGAGCTTGTTGAAGATACGCAGGCCATTGCCCGTTACGGTCGTAATGTTACGAAGATGGATGCC
TTTGGCTGTACCAGCCGGGGGCAGGCACACCGCGCCGGGCTGTGGCTGATTAAAACAGAACTGCTGGAAACGCAGAC
CGTGGATTTCAGCGTCGGCGCAGAAGGGCTTCGCCATGTACCGGGCGATGTTATTGAAATCTGCGATGATGACTATG
CCGGTATCAGCACCGGTGGTCGTGTGCTGGCGGTGAACAGCCAGACCCGGACGCTGACGCTCGACCGTGAAATCACG
CTGCCATCCTCCGGTACCGCGCTGATAAGCCTGGTTGACGGAAGTGGCAATCCGGTCAGCGTGGAGGTTCAGTCCGT
CACCGACGGCGTGAAGGTAAAAGTGAGCCGTGTTCCTGACGGTGTTGCTGAATACAGCGTATGGGAGCTGAAGCTGC
CGACGCTGCGCCAGCGACTGTTCCGCTGCGTGAGTATCCGTGAGAACGACGACGGCACGTATGCCATCACCGCCGTG
CAGCATGTGCCGGAAAAAGAGGCCATCGTGGATAACGGGGCGCACTTTGACGGCGAACAGAGTGGCACGGTGAATGG
TGTCACGCCGCCAGCGGTGCAGCACCTGACCGCAGAAGTCACTGCAGACAGCGGGGAATATCAGGTGCTGGCGCGAT
GGGACACACCGAAGGTGGTGAAGGGCGTGAGTTTCCTGCTCCGTCTGACCGTAACAGCGGACGACGGCAGTGAGCGG
CTGGTCAGCACGGCCCGGACGACGGAAACCACATACCGCTTCACGCAACTGGCGCTGGGGAACTACAGGCTGACAGT
CCGGGCGGTAAATGCGTGGGGGCAGCAGGGCGATCCGGCGTCGGTATCGTTCCGGATTGCCGCACCGGCAGCACCGT
CGAGGATTGAGCTGACGCCGGGCTATTTTCAGATAACCGCCACGCCGCATCTTGCCGTTTATGACCCGACGGTACAG
TTTGAGTTCTGGTTCTCGGAAAAGCAGATTGCGGATATCAGACAGGTTGAAACCAGCACGCGTTATCTTGGTACGGC
GCTGTACTGGATAGCCGCCAGTATCAATATCAAACCGGGCCATGATTATTACTTTTATATCCGCAGTGTGAACACCG
TTGGCAAATCGGCATTCGTGGAGGCCGTCGGTCGGGCGAGCGATGATGCGGAAGGTTACCTGGATTTTTTCAAAGGC
AAGATAACCGAATCCCATCTCGGCAAGGAGCTGCTGGAAAAAGTCGAGCTGACGGAGGATAACGCCAGCAGACTGGA
GGAGTTTTCGAAAGAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGCAGACCAAAG
ACGGCAAACATTATGTCGCGGGTATTGGCCTCAGCATGGAGGACACGGAGGAAGGCAAACTGAGCCAGTTTCTGGTT
GCCGCCAATCGTATCGCATTTATTGACCCGGCAAACGGGAATGAAACGCCGATGTTTGTGGCGCAGGGCAACCAGAT
ATTCATGAACGACGTGTTCCTGAAGCGCCTGACGGCCCCCACCATTACCAGCGGCGGCAATCCTCCGGCCTTTTCCC
TGACACCGGACGGAAAGCTGACCGCTAAAAATGCGGATATCAGTGGCAGTGTGAATGCGAACTCCGGGACGCTCAGT
AATGTGACGATAGCTGAAAACTGTACGATAAACGGTACGCTGAGGGCGGAAAAAATCGTCGGGGACATTGTAAAGGC
GGCGAGCGCGGCTTTTCCGCGCCAGCGTGAAAGCAGTGTGGACTGGCCGTCAGGTACCCGTACTGTCACCGTGACCG
ATGACCATCCTTTTGATCGCCAGATAGTGGTGCTTCCGCTGACGTTTCGCGGAAGTAAGCGTACTGTCAGCGGCAGG
ACAACGTATTCGATGTGTTATCTGAAAGTACTGATGAACGGTGCGGTGATTTATGATGGCGCGGCGAACGAGGCGGT
ACAGGTGTTCTCCCGTATTGTTGACATGCCAGCGGGTCGGGGAAACGTGATCCTGACGTTCACGCTTACGTCCACAC
GGCATTCGGCAGATATTCCGCCGTATACGTTTGCCAGCGATGTGCAGGTTATGGTGATTAAGAAACAGGCGCTGGGC
ATCAGCGTGGTCTGAGTGTGTTACAGAGGTTCGTCCGGGAACGGGCGTTTTATTATAAAACAGTGAGAGGTGAACGA
TGCGTAATGTGTGTATTGCCGTTGCTGTCTTTGCCGCACTTGCGGTGACAGTCACTCCGGCCCGTGCGGAAGGTGGA
CATGGTACGTTTACGGTGGGCTATTTTCAAGTGAAACCGGGTACATTGCCGTCGTTGTCGGGCGGGGATACCGGTGT
GAGTCATCTGAAAGGGATTAACGTGAAGTACCGTTATGAGCTGACGGACAGTGTGGGGGTGATGGCTTCCCTGGGGT
TCGCCGCGTCGAAAAAGAGCAGCACAGTGATGACCGGGGAGGATACGTTTCACTATGAGAGCCTGCGTGGACGTTAT
GTGAGCGTGATGGCCGGACCGGTTTTACAAATCAGTAAGCAGGTCAGTGCGTACGCCATGGCCGGAGTGGCTCACAG
TCGGTGGTCCGGCAGTACAATGGATTACCGTAAGACGGAAATCACTCCCGGGTATATGAAAGAGACGACCACTGCCA
GGGACGAAAGTGCAATGCGGCATACCTCAGTGGCGTGGAGTGCAGGTATACAGATTAATCCGGCAGCGTCCGTCGTT
GTTGATATTGCTTATGAAGGCTCCGGCAGTGGCGACTGGCGTACTGACGGATTCATCGTTGGGGTCGGTTATAAATT
CTGATTAGCCAGGTAACACAGTGTTATGACAGCCCGCCGGAACCGGTGGGCTTTTTTGTGGGGTGAATATGGCAGTA
AAGATTTCAGGAGTCCTGAAAGACGGCACAGGAAAACCGGTACAGAACTGCACCATTCAGCTGAAAGCCAGACGTAA
CAGCACCACGGTGGTGGTGAACACGGTGGGCTCAGAGAATCCGGATGAAGCCGGGCGTTACAGCATGGATGTGGAGT
ACGGTCAGTACAGTGTCATCCTGCAGGTTGACGGTTTTCCACCATCGCACGCCGGGACCATCACCGTGTATGAAGAT
TCACAACCGGGGACGCTGAATGATTTTCTCTGTGCCATGACGGAGGATGATGCCCGGCCGGAGGTGCTGCGTCGTCT
TGAACTGATGGTGGAAGAGGTGGCGCGTAACGCGTCCGTGGTGGCACAGAGTACGGCAGACGCGAAGAAATCAGCCG
GCGATGCCAGTGCATCAGCTGCTCAGGTCGCGGCCCTTGTGACTGATGCAACTGACTCAGCACGCGCCGCCAGCACG
TCCGCCGGACAGGCTGCATCGTCAGCTCAGGAAGCGTCCTCCGGCGCAGAAGCGGCATCAGCAAAGGCCACTGAAGC
GGAAAAAAGTGCCGCAGCCGCAGAGTCCTCAAAAAACGCGGCGGCCACCAGTGCCGGTGCGGCGAAAACGTCAGAAA
CGAATGCTGCAGCGTCACAACAATCAGCCGCCACGTCTGCCTCCACCGCGGCCACGAAAGCGTCAGAGGCCGCCACT
TCAGCACGAGATGCGGTGGCCTCAAAAGAGGCAGCAAAATCATCAGAAACGAACGCATCATCAAGTGCCGGTCGTGC
AGCTTCCTCGGCAACGGCGGCAGAAAATTCTGCCAGGGCGGCAAAAACGTCCGAGACGAATGCCAGGTCATCTGAAA
CAGCAGCGGAACGGAGCGCCTCTGCCGCGGCAGACGCAAAAACAGCGGCGGCGGGGAGTGCGTCAACGGCATCCACG
AAGGCGACAGAGGCTGCGGGAAGTGCGGTATCAGCATCGCAGAGCAAAAGTGCGGCAGAAGCGGCGGCAATACGTGC
AAAAAATTCGGCAAAACGTGCAGAAGATATAGCTTCAGCTGTCGCGCTTGAGGATGCGGACACAACGAGAAAGGGGA
TAGTGCAGCTCAGCAGTGCAACCAACAGCACGTCTGAAACGCTTGCTGCAACGCCAAAGGCGGTTAAGGTGGTAATG
GATGAAACGAACAGAAAAGCCCACTGGACAGTCCGGCACTGACCGGAACGCCAACAGCACCAACCGCGCTCAGGGGA

ACAAACAATACCCAGATTGCGAACACCGCTTTTGTACTGGCCGCGATTGCAGATGTTATCGACGCGTCACCTGACGC
ACTGAATACGCTGAATGAACTGGCCGCAGCGCTCGGGAATGATCCAGATTTTGCTACCACCATGACTAACGCGCTTG
CGGGTAAACAACCGAAGAATGCGACACTGACGGCGCTGGCAGGGCTTTCCACGGCGAAAAATAAATTACCGTATTTT
GCGGAAAATGATGCCGCCAGCCTGACTGAACTGACTCAGGTTGGCAGGGATATTCTGGCAAAAAATTCCGTTGCAGA
TGTTCTTGAATACCTTGGGGCCGGTGAGAATTCGGCCTTTCCGGCAGGTGCGCCGATCCCGTGGCCATCAGATATCG
TTCCGTCTGGCTACGTCCTGATGCAGGGGCAGGCGTTTGACAAATCAGCCTACCCAAAACTTGCTGTCGCGTATCCA
TCGGGTGTGCTTCCTGATATGCGAGGCTGGACAATCAAGGGGAAACCCGCCAGCGGTCGTGCTGTATTGTCTCAGGA
ACAGGATGGAATTAAGTCGCACACCCACAGTGCCAGTGCATCCGGTACGGATTTGGGGACGAAAACCACATCGTCGT
TTGATTACGGGACGAAAACAACAGGCAGTTTCGATTACGGCACCAAATCGACGAATAACACGGGGGCTCATGCTCAC
AGTCTGAGCGGTTCAACAGGGGCCGCGGGTGCTCATGCCCACACAAGTGGTTTAAGGATGAACAGTTCTGGCTGGAG
TCAGTATGGAACAGCAACCATTACAGGAAGTTTATCCACAGTTAAAGGAACCAGCACACAGGGTATTGCTTATTTAT
CGAAAACGGACAGTCAGGGCAGCCACAGTCACTCATTGTCCGGTACAGCCGTGAGTGCCGGTGCACATGCGCATACA
GTTGGTATTGGTGCGCACCAGCATCCGGTTGTTATCGGTGCTCATGCCCATTCTTTCAGTATTGGTTCACACGGACA
CACCATCACCGTTAACGCTGCGGGTAACGCGGAAAACACCGTCAAAAACATTGCATTTAACTATATTGTGAGGCTTG
CATAATGGCATTCAGAATGAGTGAACAACCACGGACCATAAAAATTTATAATCTGCTGGCCGGAACTAATGAATTTA
TTGGTGAAGGTGACGCATATATTCCGCCTCATACCGGTCTGCCTGCAAACAGTACCGATATTGCACCGCCAGATATT
CCGGCTGGCTTTGTGGCTGTTTTCAACAGTGATGAGGCATCGTGGCATCTCGTTGAAGACCATCGGGGTAAAACCGT
CTATGACGTGGCTTCCGGCGACGCGTTATTTATTTCTGAACTCGGTCCGTTACCGGAAAATTTTACCTGGTTATCGC
CGGGAGGGGAATATCAGAAGTGGAACGGCACAGCCTGGGTGAAGGATACGGAAGCAGAAAAACTGTTCCGGATCCGG
GAGGCGGAAGAAACAAAAAAAAGCCTGATGCAGGTAGCCAGTGAGCATATTGCGCCGCTTCAGGATGCTGCAGATCT
GGAAATTGCAACGAAGGAAGAAACCTCGTTGCTGGAAGCCTGGAAGAAGTATCGGGTGTTGCTGAACCGTGTTGATA
CATCAACTGCACCTGATATTGAGTGGCCTGCTGTCCCTGTTATGGAGTAATCGTTTTGTGATATGCCGCAGAAACGT
TGTATGAAATAACGTTCTGCGGTTAGTTAGTATATTGTAAAGCTGAGTATTGGTTTATTTGGCGATTATTATCTTCA
GGAGAATAATGGAAGTTCTATGACTCAATTGTTCATAGTGTTTACATCACCGCCAATTGCTTTTAAGACTGAACGCA
TGAAATATGGTTTTTCGTCATGTTTTGAGTCTGCTGTTGATATTTCTAAAGTCGGTTTTTTTTCTTCGTTTTCTCTA
ACTATTTTCCATGAAATACATTTTTGATTATTATTTGAATCAATTCCAATTACCTGAAGTCTTTCATCTATAATTGG
CATTGTATGTATTGGTTTATTGGAGTAGATGCTTGCTTTTCTGAGCCATAGCTCTGATATCCAAATGAAGCCATAGG
CATTTGTTATTTTGGCTCTGTCAGCTGCATAACGCCAAAAAATATATTTATCTGCTTGATCTTCAAATGTTGTATTG
ATTAAATCAATTGGATGGAATTGTTTATCATAAAAAATTAATGTTTGAATGTGATAACCGTCCTTTAAAAAAGTCGT
TTCTGCAAGCTTGGCTGTATAGTCAACTAACTCTTCTGTCGAAGTGATATTTTTAGGCTTATCTACCAGTTTTAGAC
GCTCTTTAATATCTTCAGGAATTATTTTATTGTCATATTGTATCATGCTAAATGACAATTTGCTTATGGAGTAATCT
TTTAATTTTAAATAAGTTATTCTCCTGGCTTCATCAAATAAAGAGTCGAATGATGTTGGCGAAATCACATCGTCACC
CATTGGATTGTTTATTTGTATGCCAAGAGAGTTACAGCAGTTATACATTCTGCCATAGATTATAGCTAAGGCATGTA
ATAATTCGTAATCTTTTAGCGTATTAGCGACCCATCGTCTTTCTGATTTAATAATAGATGATTCAGTTAAATATGAA
GGTAATTTCTTTTGTGCAAGTCTGACTAACTTTTTTATACCAATGTTTAACATACTTTCATTTGTAATAAACTCAAT
GTCATTTTCTTCAATGTAAGATGAAATAAGAGTAGCCTTTGCCTCGCTATACATTTCTAAATCGCCTTGTTTTTCTA
TCGTATTGCGAGAATTTTTAGCCCAAGCCATTAATGGATCATTTTTCCATTTTTCAATAACATTATTGTTATACCAA
ATGTCATATCCTATAATCTGGTTTTTGTTTTTTTGAATAATAAATGTTACTGTTCTTGCGGTTTGGAGGAATTGATT
CAAATTCAAGCGAAATAATTCAGGGTCAAAATATGTATCAATGCAGCATTTGAGCAAGTGCGATAAATCTTTAAGTC
TTCTTTCCCATGGTTTTTTAGTCATAAAACTCTCCATTTTGATAGGTTGCATGCTAGATGCTGATATATTTTAGAGG
TGATAAAATTAACTGCTTAACTGTCAATGTAATACAAGTTGTTTGATCTTTGCAATGATTCTTATCAGAAACCATAT
AGTAAATTAGTTACACAGGAAATTTTTAATATTATTATTATCATTCATTATGTATTAAAATTAGAGTTGTGGCTTGG
CTCTGCTAACACGTTGCTCATAGGAGATATGGTAGAGCCGCAGACACGTCGTATGCAGGAACGTGCTGCGGCTGGCT
GGTGAACTTCCGATAGTGCGGGTGTTGAATGATTTCCAGTTGCTACCGATTTTACATATTTTTTGCATGAGAGAATT
TGTACCACCTCCCACCGACCATCTATGACTGTACGCCACTGTCCCTAGGACTGCTATGTGCCGGAGCGGACATTACA
AACGTCCTTCTCGGTGCATGCCACTGTTGCCAATGACCTGCCTAGGAATTGGTTAGCAAGTTACTACCGGATTTTGT
AAAAACAGCCCTCCTCATATAAAAAGTATTCGTTCACTTCCGATAAGCGTCGTAATTTTCTATCTTTCATCATATTC
TAGATCCCTCTGAAAAAATCTTCCGAGTTTGCTAGGCACTGATACATAACTCTTTTCCAATAATTGGGGAAGTCATT
CAAATCTATAATAGGTTTCAGATTTGCTTCAATAAATTCTGACTGTAGCTGCTGAAACGTTGCGGTTGAACTATATT
TCCTTATAACTTTTACGAAAGAGTTTCTTTGAGTAATCACTTCACTCAAGTGCTTCCCTGCCTCCAAACGATACCTG
TTAGCAATATTTAATAGCTTGAAATGATGAAGAGCTCTGTGTTTGTCTTCCTGCCTCCAGTTCGCCGGGCATTCAAC
ATAAAAACTGATAGCACCCGGAGTTCCGGAAACGAAATTTGCATATACCCATTGCTCACGAAAAAAAATGTCCTTGT
CGATATAGGGATGAATCGCTTGGTGTACCTCATCTACTGCGAAAACTTGACCTTTCTCTCCCATATTGCAGTCGCGG
CACGATGGAACTAAATTAATAGGCATCACCGAAAATTCAGGATAATGTGCAATAGGAAGAAAATGATCTATATTTTT
TGTCTGTCCTATATCACCACAAAATGGACATTTTTCACCTGATGAAACAAGCATGTCATCGTAATATGTTCTAGCGG
GTTTGTTTTTATCTCGGAGATTATTTTCATAAAGCTTTTCTAATTTAACCTTTGTCAGGTTACCAACTACTAAGGTT
GTAGGCTCAAGAGGGTGTGTCCTGTCGTAGGTAAATAACTGACCTGTCGAGCTTAATATTCTATATTGTTGTTCTTT

CTGCAAAAAAGTGGGGAAGTGAGTAATGAAATTATTTCTAACATTTATCTGCATCATACCTTCCGAGCATTTATTAA
GCATTTCGCTATAAGTTCTCGCTGGAAGAGGTAGTTTTTTCATTGTACTTTACCTTCATCTCTGTTCATTATCATCG
CTTTTAAAACGGTTCGACCTTCTAATCCTATCTGACCATTATAATTTTTTAGAATGGTTTCATAAGAAAGCTCTGAA
TCAACGGACTGCGATAATAAGTGGTGGTATCCAGAATTTGTCACTTCAAGTAAAAACACCTCACGAGTTAAAACACC
TAAGTTCTCACCGAATGTCTCAATATCCGGACGGATAATATTTATTGCTTCTCTTGACCGTAGGACTTTCCACATGC
AGGATTTTGGAACCTCTTGCAGTACTACTGGGGAATGAGTTGCAATTATTGCTACACCATTGCGTGCATCGAGTAAG
TCGCTTAATGTTCGTAAAAAAGCAGAGAGCAAAGGTGGATGCAGATGAACCTCTGGTTCATCGAATAAAACTAATGA
CTTTTCGCCAACGACATCTACTAATCTTGTGATAGTAAATAAAACAATTGCATGTCCAGAGCTCATTCGAAGCAGAT
ATTTCTGGATATTGTCATAAAACAATTTAGTGAATTTATCATCGTCCACTTGAATCTGTGGTTCATTACGTCTTAAC
TCTTCATATTTAGAAATGAGGCTGATGAGTTCCATATTTGAAAAGTTTTCATCACTACTTAGTTTTTTGATAGCTTC
AAGCCAGAGTTGTCTTTTTCTATCTACTCTCATACAACCAATAAATGCTGAAATGAATTCTAAGCGGAGATCGCCTA
GTGATTTTAAACTATTGCTGGCAGCATTCTTGAGTCCAATATAAAAGTATTGTGTACCTTTTGCTGGGTCAGGTTGT
TCTTTAGGAGGAGTAAAAGGATCAAATGCACTAAACGAAACTGAAACAAGCGATCGAAAATATCCCTTTGGGATTCT
TGACTCGATAAGTCTATTATTTTCAGAGAAAAAATATTCATTGTTTTCTGGGTTGGTGATTGCACCAATCATTCCAT
TCAAAATTGTTGTTTTACCACACCCATTCCGCCCGATAAAAGCATGAATGTTCGTGCTGGGCATAGAATTAACCGTC
ACCTCAAAAGGTATAGTTAAATCACTGAATCCGGGAGCACTTTTTCTATTAAATGAAAAGTGGAAATCTGACAATTC
TGGCAAACCATTTAACACACGTGCGAACTGTCCATGAATTTCTGAAAGAGTTACCCCTCTAAGTAATGAGGTGTTAA
GGACGCTTTCATTTTCAATGTCGGCTAATCGATTTGGCCATACTACTAAATCCTGAATAGCTTTAAGAAGGTTATGT
TTAAAACCATCGCTTAATTTGCTGAGATTAACATAGTAGTCAATGCTTTCACCTAAGGAAAAAAACATTTCAGGGAG
TTGACTGAATTTTTTATCTATTAATGAATAAGTGCTTACTTCTTCTTTTTGACCTACAAAACCAATTTTAACATTTC
CGATATCGCATTTTTCACCATGCTCATCAAAGACAGTAAGATAAAACATTGTAACAAAGGAATAGTCATTCCAACCA
TCTGCTCGTAGGAATGCCTTATTTTTTTCTACTGCAGGAATATACCCGCCTCTTTCAATAACACTAAACTCCAACAT
ATAGTAACCCTTAATTTTATTAAAATAACCGCAATTTATTTGGCGGCAACACAGGATCTCTCTTTTAAGTTACTCTC
TATTACATACGTTTTCCATCTAAAAATTAGTAGTATTGAACTTAACGGGGCATCGTATTGTAGTTTTCCATATTTAG
CTTTCTGCTTCCTTTTGGATAACCCACTGTTATTCATGTTGCATGGTGCACTGTTTATACCAACGATATAGTCTATT
AATGCATATATAGTATCGCCGAACGATTAGCTCTTCAGGCTTCTGAAGAAGCGTTTCAAGTACTAATAAGCCGATAG
ATAGCCACGGACTTCGTAGCCATTTTTCATAAGTGTTAACTTCCGCTCCTCGCTCATAACAGACATTCACTACAGTT
ATGGCGGAAAGGTATGCATGCTGGGTGTGGGGAAGTCGTGAAAGAAAAGAAGTCAGCTGCGTCGTTTGACATCACTG
CTATCTTCTTACTGGTTATGCAGGTCGTAGTGGGTGGCACACAAAGCTTTGCACTGGATTGCGAGGCTTTGTGCTTC
TCTGGAGTGCGACAGGTTTGATGACAAAAAATTAGCGCAAGAAGACAAAAATCACCTTGCGCTAATGCTCTGTTACA
GGTCACTAATACCATCTAAGTAGTTGATTCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTT
TTATGCAAAATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTTTTATACTAACTTGA
GCGAAACGGGAAGGTAAAAAGACAAAAAGTTGTTTTTAATACCTTTAAGTGATACCAGATGGCATTGCGCCATCTGG
C

>R6K-kan-ccdB-cI857

CCTCCCACACATAACCAGGAGGTCAGATTATGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTG
TTTGTGGATGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCT
GTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATA
TGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCC
ATTAACCTGATGTTCTGGGGAATATAACCCAGAAGCTTAGCAAAAGCTAAAACCAGGAGCTATTTAATGGCAACAGT
TAACCAGCTGGTACGCAAACCACGTGCTCGCAAAGTTGCGAAAAGCAACGTGCCTGCGCTGGAAGCATGCCCGCAAA
AACGTGGCGTATGTACTCGTGTATATACTACCACTCCTAAAAAACCGAACTCCGCGCTGCGTAAAGTATGCCGTGTT
CGTCTGACTAACGGTTTCGAAGTGACTTCCTACATCGGTGGTGAAGGTCACAACCTGCAGGAGCACTCCGTGATCCT
GATCCGTGGCGGTCGTGTTAAAGACCTCCCGGGTGTTCGTTACCACACCGTACGTGGTGCGCTTGACTGCTCCGGCG
TTAAAGACCGTAAGCAGGCTCGTTCCAAGTATGGCGTGAAGCGTCCTAAGGCTTAATGGTTCGCCCGCCTAATGAGC
GGGCTTTTTTTTGAATTCTTTTTTAATTCGATCTGAAGATCAGCAGTTCAACCTGTTGATAGTACGTACTAAGCTCT
CATGTTTCACGTACTAAGCTCTCATGTTTAACGTACTAAGCTCTCATGTTTAACGAACTAAACCCTCATGGCTAACG
TACTAAGCTCTCATGGCTAACGTACTAAGCTCTCATGTTTCACGTACTAAGCTCTCATGTTTGAACAATAAAATTAA
TATAAATCAGCAACTTAAATAGCCTCTAAGGTTTTAAGTTTTATAAGAAAAAAAAGAATATATAAGGCTTTTAAAGC
TTTTAAGGTTTAACGGTTGTGGACAACAAGCCAGGGATGTAACGCACTGAGAAGCCCTTAGAGCCTCTCAAAGCAAT
TTTGAGTGACACAGGAACACTTAACGGCTGACATGGGATCCCCCTCATCAGTGCCAACATAGTAAGCCAGTATACAC
TCCGCTAGCTAACGTATGAGCACAAAAAAGAAACCATTAACACAAGAGCAGCTTGAGGACGCACGTCGCCTTAAAGC
AATTTATGAAAAAAAGAAAAATGAACTTGGCTTATCCCAGGAATCTGTCGCAGACAAGATGGGGATGGGGCAGTCAG
GCGTTGGTGCTTTATTTAATGGCATCAATGCATTAAATGCTTATAACGCCGCATTGCTTACAAAAATTCTCAAAGTT
AGCGTTGAAGAATTTAGCCCTTCAATCGCCAGAGAAATCTACGAGATGTATGAAGCGGTTAGTATGCAGCCGTCACT
TAGAAGTGAGTATGAGTACCCTGTTTTTTCTCATGTTCAGGCAGGGATGTTCTCACCTGAGCTTAGAACCTTTACCA

AAGGTGATGCGGAGAGATGGGTAAGCACAACCAAAAAAGCCAGTGATTCTGCATTCTGGCTTGAGGTTGAAGGTAAT
TCCATGACCGCACCAACAGGCTCCAAGCCAAGCTTTCCTGACGGAATGTTAATTCTCGTTGACCCTGAGCAGGCTGT
TGAGCCAGGTGATTTCTGCATAGCCAGACTTGGGGGTGATGAGTTTACCTTCAAGAAACTGATCAGGGATAGCGGTC
AGGTGTTTTTACAACCACTAAACCCACAGTACCCAATGATCCCATGCAATGAGAGTTGTTCCGTTGTGGGGAAAGTT
ATCGCTAGTCAGTGGCCTGAAGAGACGTTTGGCTGACTCGAGTTTCGACCTGCAGCCTGTTGACAATTAATCATCGG
CATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGGATCGGCCATTGAACAAGATGGATTG
CACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGA
TGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATG
AACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTC
ACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGC
CGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACC
AAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAG
CATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTCGTCGTGAC
CCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGG
GTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGAC
CGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTT
CTGAGGGGATCAATTCTCTAGAGCTCGCTGATCAGCCTCGACTGTACCGTTAGC

>recA1202 cosmid

TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
AAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCA
CCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCA
ACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT
GGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGT
TGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAAGCA
TGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTTCACAAC
ATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACAC
CACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAAC
AATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATT
GCTGATAAATCTGGAGCCGGTGAGCGTGGGTCCCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCG
TATCGTAGTTATCTACACGACGGGCAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCT
CACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAA
TTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTG
AGCGTCAGACCCCGTAGCTCGAGCCCCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG
ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGAT
CAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTA
GCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG
CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCG
GGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAG
AGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT
GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGC
TTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTT
ATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAA
ATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAAA
TTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGTGTGACAATCTAA
AAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGTAAAAATAGCCCG
CGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATGCTGTATCTGTT
CGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAATATG
CTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAA
GTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCAAACGTGC
GCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACC
GGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGT
CAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTCA
AAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAA
TGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCC

ATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGG
GTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGA
AGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTA
GTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTG
GAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCC
TCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGTGACTGAGGTATGTGCTC
TTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGATTTTGTTGTT
GCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCATGGA
AACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCC
CGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATC
AGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTA
TACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGG
TGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTG
AAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGA
TCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGC
CCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAA
TTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTCATGACTATGA
TGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGATGTGCTGATTG
TTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAAC
GTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGTAATGGCTCTCAGTCCCC
GTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTG
GTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAAT
GCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATG
AAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGC
CCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTC
GGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAA
CGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACC
GGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCG
AAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTAT
CGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGC
GCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTG
TTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTACAGATAAAGAG
GAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGCTGAAGAAGTTAT
CACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCTG
GAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGAG
AAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCACGTTTTAGTCTACGTTTATCTGTC
TTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTG
TATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGG
TCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTG
GGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAA
GTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTG
TGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTT
AACCGGGCCTCCACGCACGTTGTGATATGTAGATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACA
GGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCG
TCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGG
CGGCCGCAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTAATTTGATGCCTGGCAGTTT
ATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCGCAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTA
CTCAGGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTG
ATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCCACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTC
GGCATGGGGTCAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGA
TTTAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTTTAACGGATTCACCCTCGAGAGCT
TAAAAATCTTCGTTAGTTTCTGCTACGCCTTCGCTATCATCTACAGAGAAATCCGGCGTTGAGTTCGGGTTGCTCAG
CAGCAACTCACGTACTTTCTTCTCGATCTCTTTCGCGGTTTCCGGGTTATCTTTCAGCCAGGCAGTCGCATTCGCTT
TACCCTGACCGATCTTCTCACCTTTGTAGCTGTACCACGCGCCTGCTTTCTCGATCAGCTTCTCTTTTACGCCCAGG
TCAACCAGTTCGCCGTAGAAGTTGATACCTTCGCCGTAGAGGATCTGGAATTCAGCCTGTTTAAACGGCGCAGCGAT
TTTGTTCTTCACCACTTTCACGCGGGTTTCGCTACCCACCACGTTTTCGCCCTCTTTCACCGCGCCGATACGACGGA
TGTCGAGACGAACAGAGGCGTAGAATTTCAGCGCGTTACCACCGGTAGTGGTTTCCGGGTTACCGAACATCACACCA

ATTTTCATACGGATCTGGTTGATGAAGATCAGCAGCGTGTTGGACTTCTTCAGGTTACCCGCCAGCTTACGCATCGC
CTGGCTCATCATACGTGCCGCAAGGCCCATGTGAGAGTCGCCGATTTCGCCTTCGATTTCCGCTTTCGGCGTCAGTG
CCGCCACGGAGTCAACGACGATAACGTCTACTGCGCCAGAACGCGCCAGGGCGTCACAGATTTCCAGTGCCTGCTCG
CCGGTGTCCGGCTGGGAGCACAGCAGGTTGTCGATATCGACGCCCAGTTTACGTGCGTAGATTGGGTCCAGCGCGTG
TTCAGCATCGATAAACGCACAGGTTTTACCTTCACGCTGCGCTGCGGCGATCACCTGCAGCGTCAGCGTGGTTTTAC
CGGAAGATTCCGGTCCGTAGATTTCGACGATACGGCCCATCGGCAGACCACCTGCCCCAAGCGCGATATCCAGTGAA
AGCGAACCGGTAGAGATGGTTTCCACATCCATGGAACGGTCTTCACCCAGGCGCATGATGGAGCCTTTACCAAATTG
TTTCTCAATCTGGCCCAGTGCTGCCGCCAACGCTTTCTGTTTGTTTTCGTCGATAGCCATGGTTCCACCTCCTTGAG
TTGTAGCTAGCCCAAAAAAACGGGTCAACATACGAGCCGGAAGCATAAAGTGTAAAGGATCCGCCAGGTGGCACTTT
TCGGGGAAATGTGCGCGGAACCCCTATTTGTTTAT