UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Synthetic Auxotrophs with Ligand-Dependent Essential Genes for a BL21(DE3) Biosafety Strain

Permalink
https://escholarship.org/uc/item/3x87f1zr

Author
Lopez, Gabriel Alfredo

Publication Date
2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Synthetic Auxotrophs with Ligand-Dependent Essential Genes for a BL21(DE3) Biosafety Strain

By

Gabriel Alfredo Lopez

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor J. Christopher Anderson, Chair
Professor Patricia Babbitt
Professor Kathleen Ryan
Professor John Dueber

Spring 2015

Abstract

Synthetic Auxotrophs with Ligand-Dependent Essential Genes for a BL21(DE3) Biosafety Strain

by

Gabriel Alfredo Lopez

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor J. Christopher Anderson, Chair

Synthetic auxotrophs are organisms engineered to require the presence of a particular molecule for viability. We show that these organisms can be generated by engineering ligand-dependence into essential genes. We demonstrate a method for generating a Synthetic auxotroph based on a Ligand-Dependent Essential gene (SLiDE), using six essential genes as test cases: *pheS*, *dnaN*, *tyrS*, *metG*, *orn,* and *adk*. We show that a single SLiDE strain can have a 1×10^{8}-fold increase in viability when chemically complemented with the ligand benzothiazole. The optimized SLiDE engineering protocol required less than one week and $100 USD. We combined multiple SLiDE strain alleles into the industrial *E. coli* strain BL21(DE3), yielding an organism that exceeds the biosafety criteria with an escape frequency below the limit of detection of 3×10^{-11}.

For Grampy.

You are the ideal to which I aspire.

Table of Contents

List of Figures

Preface

Synthetic biology is an approach to genetic engineering that holds potential for transforming how products are manufactured and diseases are cured. The inherent strengths of biology, near limitless design complexity and self replication, draw nervous scrutiny from people concerned that engineered organisms might one day escape from the lab with unpredictable consequences. In responding to these societal concerns, synthetic biologists have explored strategies to ensure engineered organisms remain under complete control.

We hope to contribute to the dialog between the scientific community and the public at large by describing a technology that enables a simple approach to biocontainment. The new technology, a Synthetic auxotroph based on a Ligand-Dependent Essential gene (SLiDE), works on the same basic principle as a lock and key. In order for the SLiDE organisms to survive, a series of genetic locks must be opened with a chemical key. As soon as the key is removed, the genetic locks automatically close and the organism dies. Because the chemical key can only be found in the lab, organisms attempting to escape will quickly drop dead, ensuring effective biological containment.

If proven effective, such a containment scheme might allow for the more liberal deployment of engineered organisms. Because containment is intrinsic, the physical location of an engineered organism becomes irrelevant with respect to our control over the organism's fate. Organisms that have been engineered for such a form of containment might find utility as platforms for drug delivery, live vaccines, improved nitrogen fixation, or bioremediation. There is, perhaps, a certain irony in that the public's demand for the containment of engineered organisms may actually lead to their widespread environmental deployment.

In addition to its utility as part of a biocontainment strategy, SLiDE organisms have the potential to dramatically facilitate the development and optimization of new biosynthetic pathways. As synthetic biologists, we strive to create new biological function. Our methodologies fundamentally mirror how nature approaches the problem: generate diversity and screen for a desired phenotype. While generating genetic diversity is trivial, rapidly probing the biological functionalities of billions of genotypic variants has limited our ability to engineer useful new products. SLiDE organisms might offer an easy solution.

By engineering a SLiDE organism to detect a valuable chemical, the organism can be used as a sensor for engineered enzymes that produce our desired

valuable chemical. Going back to our lock and key analogy from above, we would now be using a lock to find a key. If we were handed a jar full of keys, a closed padlock, and instructions stating that the correct key was made of solid gold, then we could use the padlock to find the golden key. SLiDE strain biosensors could work in the same way, identifying valuable catalytic chemistries instead of golden keys, and able to test ten billion “golden keys” simultaneously.

But for all of the potential applications of our technology, we feel that its greatest strength lies in its simplicity. We present a simple framework, using common tools and techniques of microbial genetics, some dating back to the early 1940’s. In striving for simplicity, we developed methods enabling the generation SLiDE strains in under one week and for less than $100 USD. We hope that the practicality of our work finds favor with investigators who might find unexpected applications towards unexplored problems.

A Note on My Monochromatic Color Scheme

I drew heavily on the insights and techniques from the 1940's-1950's: Beadle and Tatum; Stokes, Foster, and Woodward; Mitchell and Houlahan; Joshua and Esther Lederberg; Bernard Davis. This is an homage to their work. The aesthetic is intended to honor their ability to have elegantly deduced so much with such simplicity.

Acknowledgements

Professor Christopher Anderson. You brought out the best in me. My doctoral work has been one of the most challenging experiences in my life. Your uncompromising standards forced me to marshal my intellectual and emotional resources so that I might discover strengths I did not know I possessed.

Grammy. You always have a different perspective. One day I will convince you I am right.

Papa. You set me on a course of scientific discovery. I hope I have made you proud.

Mom. Three decades of unwavering love and support have kept me strong through thick and thin.

Introduction

Engineering ligand-dependence into proteins has many applications. One such application is developing genetically encoded biosensors(1–3). Such biosensors might find use as feedback control elements within engineered pathways, as metabolite sensors for engineering recombinant biosynthesis pathways, or as detectors of environmental contaminants. Engineered ligand-dependence might also be used to create synthetic auxotrophs: organisms specifically engineered to depend on a particular molecule for viability. Such organisms could be deployed as part of a biocontainment strategy, preventing the escape of genetically modified organisms from the lab or providing a means for maintaining control over organisms used in open system applications(4).

To generate such organisms, the engineering challenge of reliably creating ligand-dependent proteins must be addressed. Creating ligand-dependent phenotypes in proteins that are not normally ligand-dependent requires protein engineering. The approach to protein engineering can be broken into two smaller problems: formulating a suitable design strategy and developing an effective screening or selection methodology.

There are two principle approaches to protein engineering: rational protein engineering and directed evolution. While the specific techniques applied differ greatly between these approaches, both rational and directed evolution rely on the same basic evolutionary principles found in natural selection: generation of diversity followed by phenotypic selection.

Directed evolution more obviously resembles natural selection. Genetic diversity is created and a screen or selection serves as a form of selective pressure. In the ideal case, desired mutants have increased fitness due to the artificial selective pressure and can thus be enriched. Multiple cycles of directed evolution can be used to improve properties of engineered proteins(5, 6).

Rational protein engineering bears only conceptual similarity to natural selection. Computationally generated diversity is subjected to mathematically-based fitness algorithms to enrich (select) for desirable variants(7). Despite their conceptual similarities, these two approaches suffer from different limitations.

In rational protein engineering, algorithmic complexity and accuracy is limited by computation capacity. More accurate simulations, requiring greater computational power, can assess fewer protein variants over a given time. A

perfectly accurate folding algorithm would be useless if it took 100 years to generate a single prediction. Therefore, computational accuracy must be weighed against the computational power available to screen reasonably-sized sets of structural diversity.

In directed evolution approaches, libraries are screened or selected directly for a desired phenotype. This is a powerful approach when the desired phenotype has an easy screen or selection, like fluorescence or antibiotic resistance. Unfortunately, this is often not the case. If rational protein engineering is generally limited by scaling, directed evolution is limited by generalizability(5–7).

Semi-rational protein engineering might be considered a third approach. A hybrid methodology, it combines the strengths of rational design with directed evolution. By rationally targeting mutagenesis, diversity can be focused towards residues that are known or suspected to yield desired activity(8, 9). Screens and selections (if available) allow for targeted libraries to be quickly probed for desired activity. This translates to a more efficient search of DNA sequence space. Because of its flexibility and potential efficiency, we opted for a semi-rational protein engineering approach.

Because semi-rational protein engineering involves targeting diversity to specific parts of a protein, decisions about how to target mutagenesis had to be considered. The approach to library design would prove a central challenge of engineering ligand-dependence. The objective of developing and validating approaches to library design that reliably contained ligand-dependent phenotypes was hampered by the fundamental weakness of directed evolution: a general lack of selections and screens.

Without methodologies to isolate desirable mutants from a library background, library design approaches could not be validated(6). Therefore, the fundamental challenge of efficiently probing a high diversity library for a desired ligand-dependent phenotype had to be addressed before potential library design approaches might be validated.

This presented a conundrum. To validate our library design approach, we needed a reliable screen or selection for ligand dependence, but to validate out screens or selections needed a method for generating libraries containing ligand dependent mutants. To address this dilemma, we made an assumption. We assumed (based on experiments to be discussed) that we could devise a reasonable library design approach that would contain ligand dependent

mutants. We wagered that we could use these libraries to validate selection and screening approaches. This boot-strapping approach might then allow us to refine our library design approach, which would enable us to further polish off our selections and screens. This process could repeat until we were satisfied with both.

Because we solved our library design challenge with an assumption, the next challenge was to devise a selection or screening strategy to start the boot-strapping process. While developing an experimental framework for the selection and screening aspect of our protein engineering approach, we found it useful to reexamine the foundational methodologies and observations of genetics and molecular biology. By redefining our problem of engineering ligand-dependence more clearly as an attempt to create ligand-dependent synthetic auxotrophs, our engineering challenge matched a scientific framework established 75 years ago.

Conditional mutants were key in establishing the one gene-one enzyme hypothesis(10). They also served as an early hint of how synthetic control might be exerted over enzyme function and ultimately cellular viability. Beadle and Tatum used X-ray mutagenesis to create libraries on *Neurospora*, common bread mold. Individual mutants from the resulting pool of diversity were separated and the monoclonal isolates were grown up in test tubes containing a complete media. These mutants were then used to inoculate fresh test tubes containing minimal media (lacking intermediate metabolites) as well as test tubes containing the same type of complete media that the mutant was originally grown in. Beadle and Tatum were looking for growth defects between the two conditions(10, 11). One of their samples would prove to be a scientific jackpot.

Pyridoxineless Mutant No. 299, was the first characterized auxotroph. It happily grew in the complete media test tube, but no growth was observed in the minimal media test tube. As its name would suggest, a search for the hypothesized biochemical deficiency revealed pyridoxine's ability to complement Mutant No. 299's growth in the minimal media(11). It was the serendipitous observation of this mutant's conditional phenotype, however, that illustrated how protein function and cellular viability might be controlled.

A group working at Merck observed that Mutant No. 299's pyridoxine auxotrophy disappeared when the pH was raised from 5 to 6 (12). This result precipitated a storm of speculation as to other types of conditionalities might be generated. The first temperature-sensitive *Neurospora* mutant was isolated three years later(13).

Soon after, these findings were reproduced in bacteria(14) and eventually extended to other microorganisms.

Temperature sensitive mutants (along with other conditional phenotypes such as pH or osmolality sensitive mutants) were key in shedding light on the nature of genetics, biochemistry, and molecular biology. Most importantly, they provided evidence on the relationship between genes and enzymes, genotypes and phenotypes. However, from an engineering standpoint, we saw the conditional mutant as a model phenotype that could inform our conceptual approach for generating ligand-dependent control over protein function. By framing our engineering challenge of generating ligand-dependence more concretely as the attempt to create ligand-dependent conditional mutants (synthetic auxotrophs), then our experimental approach could benefit from established approaches used to isolate traditional metabolic auxotrophs and conditional mutants.

Over the next seven decades, scientists worked to not only characterize the molecular mechanisms of conditional mutants, but also to purposefully engineer control over protein function. Early attempts at engineering allosteric control exploited the structure-function relationship of enzyme activity. By inserting a ligand binding domain into a flexible loop of the target protein, conformational changes in the receptor domain mediated by ligand binding might extend into the catalytic domain, causing changes in tertiary structure sufficient to alter activity (15, 16). While engineering conditional protein function through domain insertion worked, the resulting phenotypes were weak (between 2-3 fold). In addition, the design and fabrication of variants were significant experimental undertakings.

A much simpler approach was elegantly demonstrated by the Karanicolas lab. They generated an artificial allosteric pocket directly into a native protein. Karanicolas showed that mutating a tryptophan to a glycine could tie enzymatic function to exogenously supplied indole. This was demonstrated in both β-glycosidase and β-glucuronidase. The elimination of a large, “buttressing” indole side-chain was hypothesized to cause the collapse of structure required for catalytic activity. This demonstration of engineered ligand-dependency illustrated a potential method for controlling an organism’s viability if applied to essential genes. However, an engineering approach for reliably extending such phenotypes to other enzymes remained elusive.

While the Karanicolas lab took a rational approach to protein engineering, screening large, multi-site combinatorial libraries were proven effective for identifying dramatic changes in protein phenotype. Sometimes the phenotypic

changes required several simultaneous mutations. Tang and Cirino demonstrated the value of a large library by altering the substrate specificity of araC. Using multi-site-saturation-mutagenesis of the ligand-binding residues in araC followed by FACS-based screening, they generated a mutant responsive to mevalonate instead of arabinose(17). The results from these different approaches suggest that many engineering targets and strategies exist for generating ligand dependence within a given protein. Unfortunately, the lack of a robust and generalizable screening or selection methodology hampers efforts towards creating synthetic ligand-dependence in genes lacking transcriptional or colorimetric outputs.

The strategies for engineering ligand-dependent protein function discussed above suggest a possible solution for synthetic auxotrophy based on applying ligand-dependent control of enzyme function towards essential genes. Such a strategy would have to overcome the fundamental challenge of reliably generating ligand dependence. In contrast to small molecule control of protein function, a different approach to synthetic auxotrophy was recently demonstrated as part of a biocontainment strategy(18, 19). Using the spare codon of a genetically recoded *E. coli*, translational control over various essential genes was imposed by introducing synthetic metabolic requirements for a non-standard amino acid (NSAA). The resulting organisms were unable to survive unless supplied with the NSAA; biocontainment was enhanced by combining multiple dependencies. Currently, this strategy is confined to organisms whose genomes have been recoded, but this might be overcome by increasing the number of available codons with an expanded genetic alphabet.

In this work we demonstrate a method for engineering a new class of conditional mutant: Synthetic auxotroph based on a Ligand-Dependent Essential gene (SLiDE). Our methodology builds upon the mechanisms underlying conditional viability observed by Beadle and Tatum. Using the Keio collection(20) to identify essential gene targets in *E. coli*, we aimed to generate de novo ligand control into native proteins, as demonstrated by Karanicolas. Based on the work of Tang and Cirino, we hypothesized that larger libraries might improve the chances of generating SLiDE strains, so we simultaneously mutagenized multiple residues of our target essential genes. Extending this train of logic to the choice of complementing molecules, we elected to use a small pool of chemicals to improve the odds of finding a SLiDE strain. Exploiting the inherent selectability of essential gene phenotypes, we adapted Bernard Davis' penicillin technique(21) to efficiently identify SLiDE mutants for *pheS*, *dnaN*, *tyrS*, *metG*, *orn,* and *adk*. To demonstrate the utility of SLiDE strains for an industrial biosafety application, we

combined multiple alleles in the BL21(DE3). Combination of two SLiDE genes into the industrially relevant strain(22) BL21(DE3) created a biosafety strain with an escape frequency of 5×10^{-10}. Combination of three SLiDE genes in BL21(DE3) resulted in an escape frequency below the limit of detection of 3×10^{-11}.

SLiDE Library Design

To identify SLiDE strains, we began with a candidate list of essential genes, necessary for *E. coli's* viability(20). We then identified the subset of these genes with a solved crystal structure to choose residues for mutagenesis. Mutations were targeted to regions near the surface, but still within the hydrophobic core. Portions of the crystal structure containing groups of large, hydrophobic residues (Trp, Phe, Met, Ile, Leu) were subjected to targeted mutagenesis in three ways, as follows.

In our first approach (on *pheS* and *dnaN*), mutagenesis was targeted such that a central large hydrophobic residue was mutated to glycine while surrounding residues were randomized using the degenerate codon NNK (see Figure 6A for library architecture—note the cluster of mutations on to the right. Figure 19A provides another example of this library architecture). Initially, these libraries were plasmid based (see Figure 13 and 14 for plasmid library DNA fabrication approach) and were transformed into $pheS^{ts}$ or $dnaN^{ts}$ temperature sensitive strains at the permissive temperature. For selections, screens, and phenotypic analysis, the libraries were grown at the restrictive temperature (see Figure 1C and 15 for selection strategies). This would abolish function of the conditional genomic essential gene in question, while uncovering the phenotype encoded by the plasmid-borne library member. We switched to generating libraries directly on the genomic copy of the targeted essential gene (see Figures 16, 17, and 18 for genome library fabrication approaches). This reduced experimental complexity and allowed us to engineer essential genes lacking temperature sensitive mutants.

In our second approach (on *tyrS* and *metG*), libraries were generated on the genome and targeted to similar hydrophobic domains, but mutagenesis was confined in DNA sequence such that all sites fit within a 21-bp window. This allowed a single 60-bp oligo to contain genome-targeting homology regions and mutagenic NNK codons (see Figure16 for library fabrication approach). Our preferred target for mutagenesis was a β strand passing through a hydrophobic core. In such β strands, every other amino acid will generally point in the same direction within the protein's secondary structure (Figure 19B illustrates orientation in a β strand). We targeted these β strands by randomizing a set of four amino acids on one or both sides of a β sheet (see figures 7A and 8A for library architecture).

Our third approach (on *adk* and *orn*) was similar to the second, with the additional design parameter of constraining mutagenesis to within 60 bp of the 5' end of the essential gene to be engineered. This simplified library fabrication, because a single PCR (using a selectable marker as template) could generate a genomic integration fragment. The integration cassette consisted of an antibiotic marker, a recoded 5' sequence of the targeted gene (to prevent premature crossover), and amino acid degeneracies (see Figure 17 for library fabrication approach).

In addition to our three targeted mutagenesis methodologies, random mutagenesis was also explored as a means of improving weak, initial SLiDE phenotypes. Error prone PCR was used to generate libraries on SLiDE alleles that were used in both plasmid and genome library directed evolution experiments (see Figures 14 and 17 for random mutagenesis approaches).

SLiDE Strain Engineering Approach

Naïve essential gene libraries contain three basic phenotypes: viable mutants (WT-like activity), lethal mutants, and SLiDE mutants. In order to isolate SLiDE strains, essential gene libraries were passed through a dual selection consisting of a positive selection based on chemically-complemented growth and a negative selection based on Bernard Davis' penicillin technique(21). Survivors of the dual selection were screened for the desired phenotype(23), using a automated colony picking and pin tool replica gridding.

For the positive selection, essential genes library strains were grown in the presence of a single chemical or a pool of exogenously supplied small molecules (Figure 2A). We used a pool of ligands to increase the efficiency of the selection process. A four chemical pool allowed us to run a single selection instead of four individual selections for each chemical(15). Under these permissive conditions, lethal mutants would fail to propagate, viable mutants would propagate, and any ligand-dependent mutants would be viable through chemical complementation (Figure 1C - Positive Selection). Because lethal mutants were not viable, the essential gene library would then contain only viable strains and SLiDE strains. Library members surviving the positive selection were subjected to a penicillin technique negative selection.

The penicillin technique negative selection relies on penicillin's growth-based mode of action. For the negative selection, surviving library members were transferred to a restrictive growth condition by removing the complementing ligands. The library was allowed to grow for 1-2 hours to account for phenotypic lag. Penicillin was then added to the growth culture. In the absence of complementing ligands, SLiDE strains were unable to grow and remain unharmed by penicillin. However, viable library members (not requiring chemical for growth) continued to divide in the restrictive condition and were subject to penicillin's antibiotic activity (Figure 1C – Negative Selection)(21).

A second positive selection was performed in order to enrich for SLiDE strains. Library members surviving the negative selection were collected, residual penicillin was washed away, and cells were spread on LB agar plates containing complementing chemicals (Figure 1C – Enrichment). Any survivors were screened for ligand-dependent growth by replica-spotting(23) on LB agar plates with and without complementing chemicals.

After the dual selection, surviving colonies were screening for SLiDE strain phenotypes. Colonies were either picked by hand, replica plated, or picked by an automated colony picking robot. The robot to increased our throughput and we used it for most of our screens. The robot picked colonies into 384 well plates which were gridded onto LB agar plates not containing any complementing chemicals and LB agar plates that did contain the complementing chemicals. The next day the plates were scanned and overlaid. Figure 1D shows the overlaid image used for screening *tyrS* SLiDE strains. The black dots indicate potential SLiDE strains.

Potential SLiDE strains spots were re-suspended, serial diluted, and the dilutions spotted onto LB agar plates containing no complementing chemicals and each of the individual chemical constituents of the chemical pool. Figure 1E shows a perliminary analysis of potential *tyrS* and *metG* SLiDE strains. Note that the serial dilutions spotted on the top plate (no chemicals) fail to grow, while the same sample grow robustly on the bottom plate (containing 1mM benzothiazole).

We chose potential complementing chemicals based on three practical criteria: low-cost, low-risk (non-explosive, non-flammable, and non-carcinogenic), and media-soluble. In addition to practical considerations, we hypothesized that side-chain-like molecules might mediate chemical complementation similar to how indole was able to complement enzyme activity in Karanicolas' experiments.

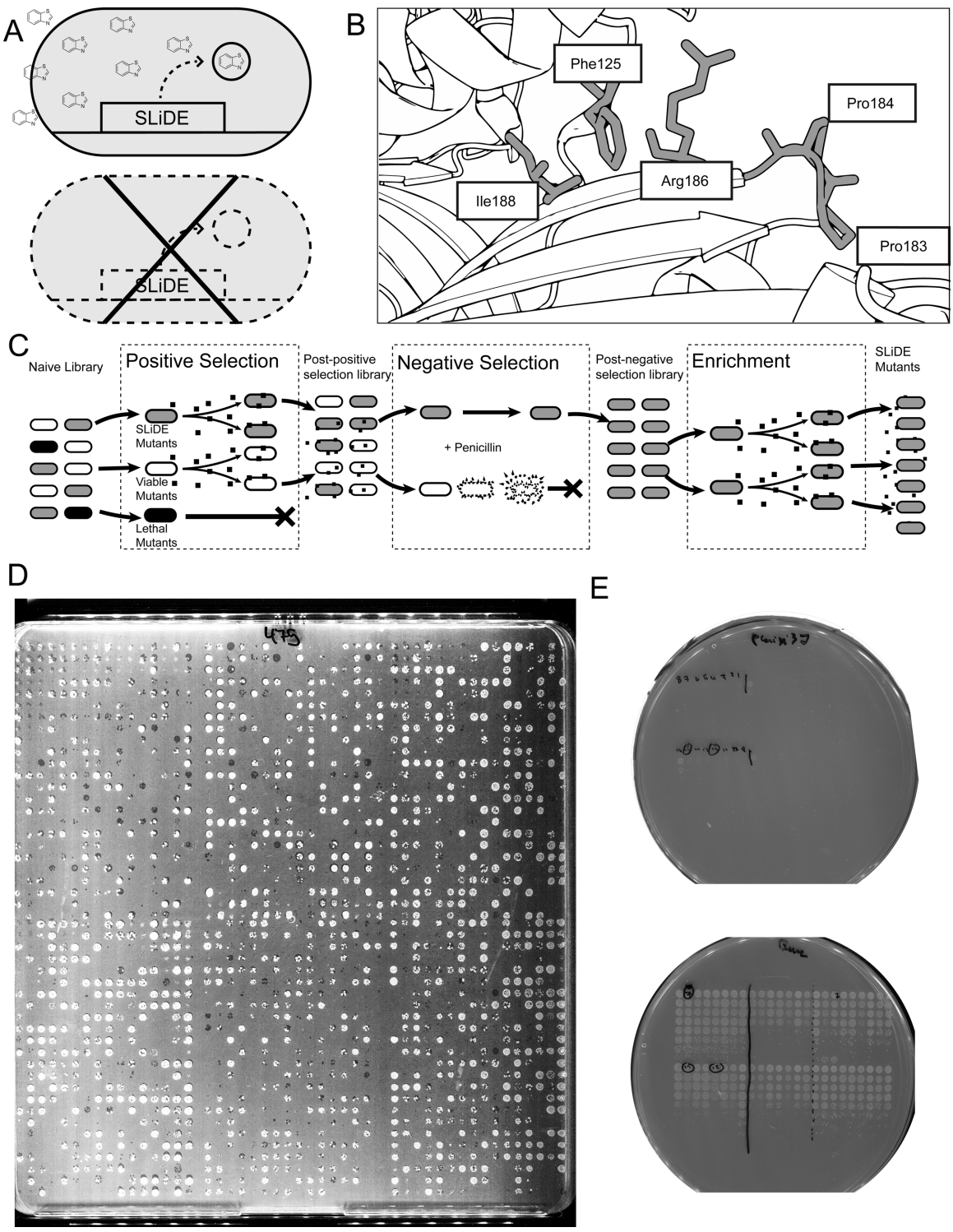

A
SLiDE
SLiDE
B
Phe125
Pro184
Arg186
Ile188
Pro183
C
Naive Library
Positive Selection
Post-positive selection library
Negative Selection
Post-negative selection library
Enrichment
SLiDE Mutants
SLiDE Mutants
Viable Mutants
Lethal Mutants
+ Penicillin
D
E

Figure 1. SLiDE Strain Engineering Strategy

A. SLiDE strains express a critical protein functional only when chemically complemented by a ligand (solid lines). In the absence of the ligand, the strain fails to propagate (dashed lines).

B. Region of *pheS* containing residues mutated in pheS.GL2 (highlighted in grey).

C. A dual selection based on chemical complementation and the penicillin technique. Black = lethal, grey = SLiDE strain, white = viable, ■ = ligand.

D. Representative replica screening image. An overlay of the restrictive and permissive screening plates for *tyrS* SLiDE strain screening. White spots grow in the absence of chemicals, while dark spots are potential SLiDE strains.

E. Preliminary analysis plates for *tyrS* and *metG* SLiDE strains (derived from screening plate to left). At top is an LB agar plate containing no chemicals. Below is a plate containing 1mM benzothiazole. Ten-fold serial dilutions were spotted onto both conditions.

SLiDE Mutant Engineering of Six Essential Genes

SLiDE strains were generated for six essential genes (*pheS*, *dnaN*, *tyrS*, *metG*, *orn* and *adk*). Each SLiDE strain carried between 3 and 7 mutations at the targeted essential gene (Table 1, Figure 2B). A mixture of 4 small molecules (Figure 2A) was used to generate SLiDE mutants. Each SLiDE strain was tested against each small molecule individually. The most promiscuous strain, metG.GL15, was complemented by all four ligands. Other mutants responded either to benzothiazole and indole or benzothiazole and 2-aminobenzothiazole. Additionally, escape frequency was measured on media lacking any of the 4 ligands (Figure 2D). The highest escape frequency was 8 x 10^{-4} for adk.GL1, while the lowest escape frequency was 3 x 10^{-9} for pheS.GL2.

Mutant	Mutations						
dnaN.GL7	H191N	R240C	I317S	F319V	L340T	V347I	S345C
pheS.GL2	F125G	P183T	P184A	R186A	I188L		
tyrS.GL7	L36V	C38A	F40G	*P42P*			
metG.GL15	E45Q	N47R	I49G	A51C			
adk.GL1	*I3I*	I4L	L5I	L6G			
orn.GL1	N7S	*W9W*	*D11D*	L13T			

Table 1. A List of Mutations for Selected SLiDE Strains

Italicized and underlined residues were mutagenized but remained wild type.

SLiDE strain pheS.GL2 was derived from a directed evolution experiment using both targeted and random mutagenesis. SLiDE strain pheS.GL2 failed to grow in the absence of benzothiazole, with an escape frequency of 3 x 10^{-9} (Figure 2D). All of the mutations in pheS.GL2 are contained within a single stretch of approximately 15 Angstroms according the crystal structure of the WT protein. In addition, all mutations reside between 15-25 Angstroms from the AMP substrate in the active site(Figure 3, 5A).

Reversion analysis of pheS.GL2 pointed to second-site genetic suppression as the primary mode of escape. Out of six escape mutants sequenced, five mutants contained Q169H, (an active site residue, approximately 10 Angstroms away

from the substrate) and one mutant contained T162N (a near-active site residue, approximately 20 Angstroms away from the AMP substrate) (Table 2, Figure 3).

Revertant	Mutations
pheS.GL2R1	T162N
pheS.GL2R2	Q169H
pheS.GL2R3	Q169H
pheS.GL2R4	Q169H
pheS.GL2R5	Q169H S95F
pheS.GL2R6	Q169H S95F T183P

Table 2. A List of Mutations in Six pheS.GL2 Escape Mutants

SLiDE strain tyrS.GL7 was generated in under two weeks, by a single step that consisted of randomizing four hydrophobic core amino acids, positive and negative genetic selection, and phenotypic screening (see Figure 16 for library fabrication approach). SLiDE strain tyrS.GL7 showed comparable escape frequency to pheS.GL2 (Figure 2D) and exhibited dose dependent growth between 250-1000 μM benzothiazole (Figure 2C). Like tyrS.GL7, SLiDE strain adk.GL1 was generated in single cycle of protein engineering, but the library was designed to reduce fabrication complexity and to drive material costs down to a minimum (see Figure 17 for library fabrication approach). This process yielded a mutant with an escape frequency of 8×10^{-4} in five days and for less than $100 USD (Figure 2D).

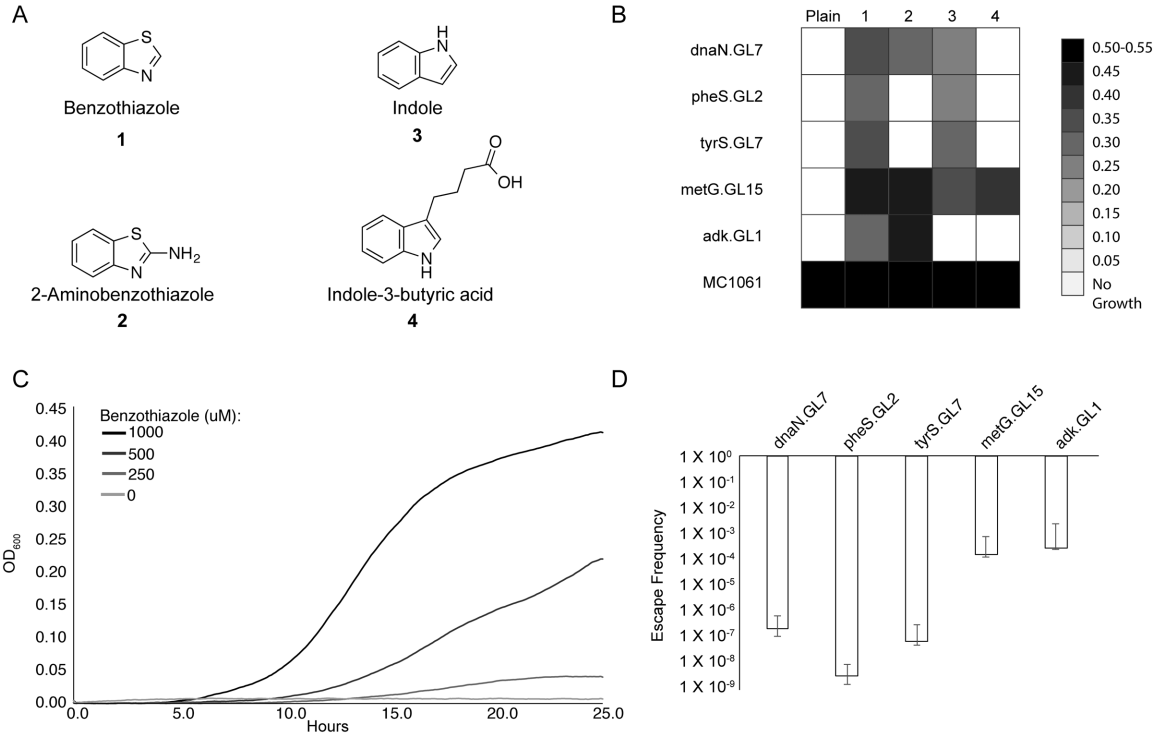

A
Benzothiazole
1
Indole
3
2-Aminobenzothiazole
2
Indole-3-butyric acid
4
B
Plain
1
2
3
4
dnaN.GL7
pheS.GL2
tyrS.GL7
metG.GL15
adk.GL1
MC1061
0.50-0.55
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
No Growth
C
Benzothiazole (uM):
1000
500
250
0
OD600
Hours
D
dnaN.GL7
pheS.GL2
tyrS.GL7
metG.GL15
adk.GL1
Escape Frequency

Figure 2. Overview of Five SLiDE Strains

A. Molecules used as a pool for SLiDE strain engineering.

B. A heat map of maximum specific growth rate for each SLiDE strain or WT MC1061 +/- each ligand. Samples that failed to grow (change in OD_{600} < 2-fold) are shown in white.

C. Representative growth curve of SLiDE strain tyrS.GL7 showing dose-dependent growth at different concentrations of benzothiazole.

D. Escape frequencies of various SLiDE strains (CFU growing on restrictive condition divided by CFU growing on permissive condition).

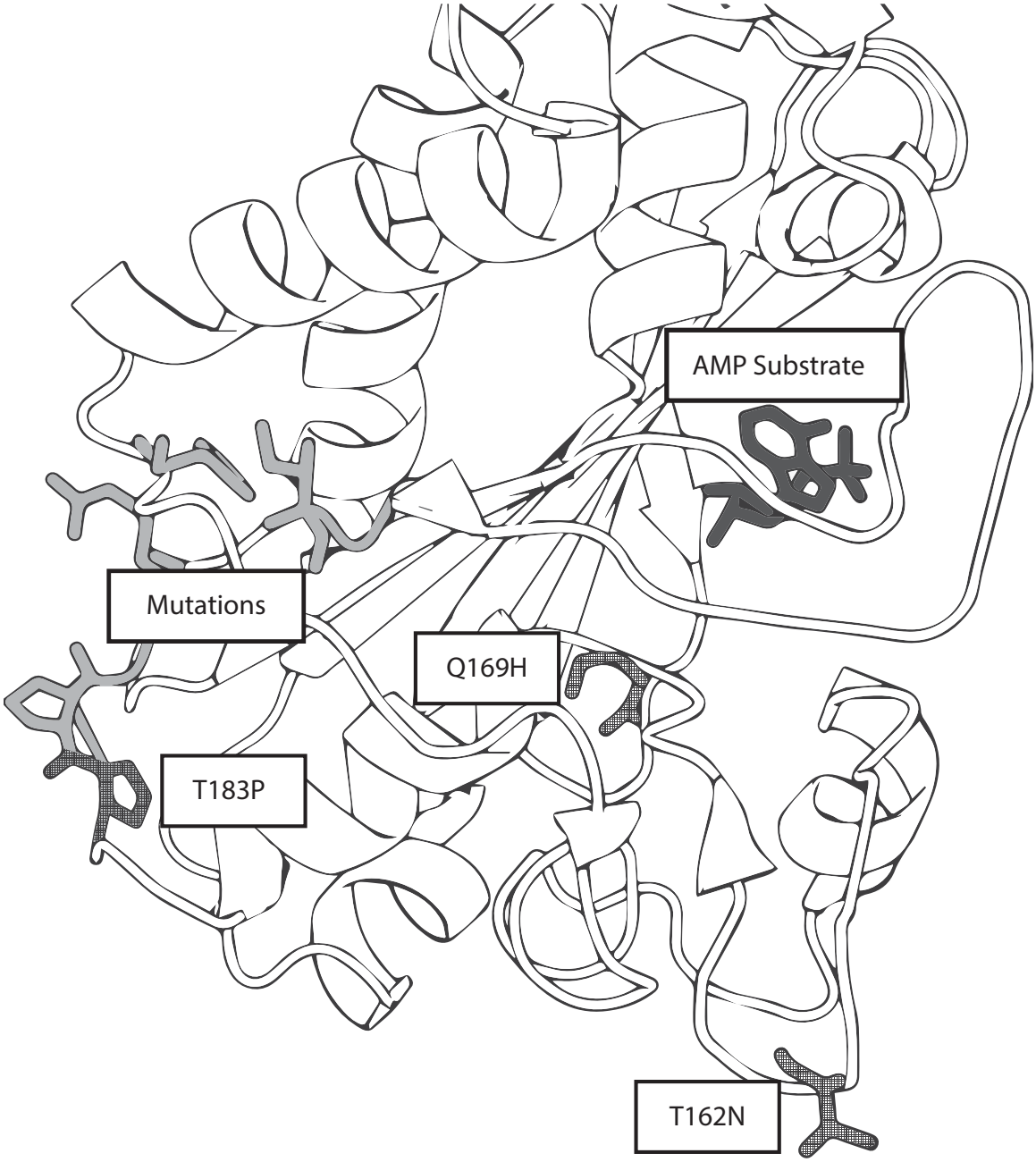

AMP Substrate
Mutations
Q169H
T183P
T162N

Figure 3. Mutations of pheS.GL2 Escape Mutants

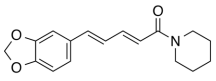
The location, with respect to the WT crystal structure, of near-active-site mutations found in pheS.GL2 escape mutants listed in Table 2. Light grey residues are library mutations, the dark grey molecule is the AMP substrate, and hatched residues are mutations discovered in escape mutants.

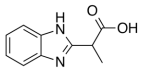
SLiDE Strain Specificity of Chemical Complementation

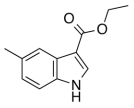
To characterize the specificity of the SLiDE strains towards ligands, we examined all strains discussed in this study for complementation by a panel of 30 additional ligands. Only two (indole-3-acetic acid and L-histidine methyl ester) were found to complement metG.GL15 (Figure 4B); the remaining 28 ligands produced no growth in the other SLiDE strains (Figure 4A).

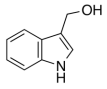
We found the lack of promiscuity to be both a relief and a disappointment. The specificity was beneficial in that SLiDE strains might be less prone to unpredicted chemical complementation from metabolic intermediates. However, the lack of promiscuity was a disappointment in that altering substrate specificity to different molecules might not be as easy as hoped. The investigation of specificity and the feasibility of altering substrate specificity are of interest for future work.

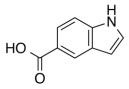
A


Piperine

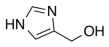

2-benzimidazole propionic acid

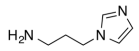

ethyl-5-methylindole-3-carboxylate


indole-3-carbinol


indole-5-carboxylic acid

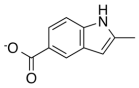

2-ethyl-2-oxazoline

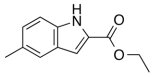

4(5)-(hydroxymethyl)-imidazole


1(3-aminopropyl) imidazole


1-2-4-triazole-3-carboxylic acid

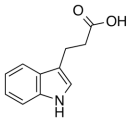

2-butanol

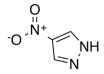

methylindole-5-carboxylate

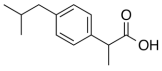

ethyl-5-methylindole-2-carboxylate


methylindole-3-carboxylate


nicotine


3-indolepropionic acid

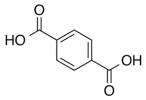

4-nitro-1H-pyrazole


ibuprofen

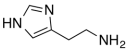

2-ethyl-2-oxazoline

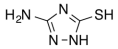

4-amino-4H-1,2,3-triazole

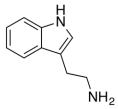

indole-3-carboxaldehyde


terephthalic acid


camphor


histamine


3-amino-1,2,4-triazole-5-thiol

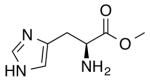

tryptamine


caprolactam


imidazole


pyrazole

B


l-histidine methyl ester

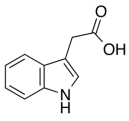

indole 3 acetic acid

Figure 4. Promiscuity Screen of SLiDE Strains

A. The best SLiDE mutants were tested for promiscuous chemical complementation by a panel of 30 ligands. Of these, 28 failed to mediate any chemical complementation.

B. Only two chemicals were able to complement metG.GL15—the most promiscuous SLiDE strain.

SLiDE Strain pheS.GL2

The essential gene *pheS* is responsible for ligation of the phenylalanine tRNA and its cognate amino acid. Without this gene, phenylalanine cannot be incorporated into growing peptide chains during translation(24).

The development of pheS.GL2 made use of every technique described in this study. We started off by generating plasmid libraries on wildtype *pheS*, transforming them into *pheS*^{ts} strains, and growing them at the permissive temperature. The *pheS* mutants transformed into *pheS*^{ts} cells will grow at the permissive condition despite a possible lethal phenotype on the plasmid because the genomic temperature sensitive gene is functional at the permissive temperature. However, when grown at the restrictive temperature, the genomic copy is non-functional and the phenotype of the plasmid-encoded mutants is revealed. (Figure 13, in the methods section, presents a visualization of library fabrication. Figure 15, also in the methods section, presents a visualization of the temperature sensitive screening approach.)

Our first libraries were conservative; we mutated a single phenylalanine or tryptophan residue to a glycine, almost exactly replicating Karanicolas' experiment(25). We found several phenylalanine to glycine mutations that resulted in a 10-fold ligand dependent phenotype, complemented by 1mM styrene(Table 3).

In order to improve the phenotype, we generated two-member targeted mutagenesis libraries, randomizing two residues that plausibly interacted with the initial phenylalanine (now a glycine) whose mutation had led to the initial phenotype. A two-residue saturation mutagenesis library contained 400 theoretical variants (20^2). Using an automated colony picker, we picked 768 variants (two 384 well plates) for each two site library in order to obtain two-fold library coverage. This was slightly lower than the recommended 3-fold coverage required to sample 95% of a library(26), but it increased throughput, allowing us to process multiple two-residue libraries in parallel.

The colonies that were picked into 384 well plates were stamped onto large screening plates and screened for no growth at the restrictive temperature and in the absence of complementing chemicals. Colonies unable to grow at the restrictive temperature were consolidated and then screened for growth at the restrictive temperature, but this time with the addition of a pool of potential complementing chemicals.

After several rounds of directed evolution using two-site targeted mutagenesis, the resulting *pheS* mutants had escape frequencies on the order of 1 x 10^{-3} (Table 3). Further rounds of directed evolution served to improve the phenotypes merely 10-fold. This was not sufficient for the intended application of biosafety (requiring escape frequencies below 1 x 10^{-8}).

Based on the steady, but very slow improvement in ligand-dependent phenotypes (as determined by reductions in escape frequency) we hypothesized that we were under-sampling protein sequence space. This was a result of our fundamental limitation of throughput. If we could inspect libraries containing millions or billions of variants, we were confident that we could find much improved mutants with far lower escape frequencies. Therefore, we concluded the key to improving our phenotypes required a dramatic increase in throughput.

Our desired approach to increasing throughput was simple: filter out useless sequence diversity. This would allow us to focus on more interesting mutants. Filtering out useless diversity was easy to do with lethal mutants, because they simply didn't grow. If an analogous approach could be employed for filtering wildtype-like viable mutants, then large libraries could be reduced to a manageable size.

If a large sequence space could be "pre-enriched" for ligand-dependent phenotypes, then our existing screening approaches could identify mutants with improved characteristics. It was at this point that the classical methodologies proved so valuable. By framing the problem of generating ligand-dependent proteins in the context of early experiments involving conditional mutants, we could use proven methodologies to our advantage. A particularly powerful approach to generating auxotrophs and conditional mutants was Bernard Davis' penicillin technique(21) (Figure 1C).

Unfortunately, the first several attempts of the penicillin technique on *pheS* libraries failed. However (due to stubbornness), the methodology was demonstrated to effectively enrich for SLiDE alleles with *dnaN* plasmid libraries (discussed next chapter). Despite continued attempts, the penicillin technique refused to improve *pheS*-based SLiDE alleles with plasmid-based libraries. We speculate that the penicillin technique may have been more effective with *dnaN* plasmid libraries because of its bactericidal phenotype. This makes it much harder for escape mutants to contaminate the resulting negative selection. In contrast, bacteriostatic *pheS* mutants are never fully dead. This allows more time

for desirable mutants to be unintentionally killed or escape mutants to contaminate the library after selection.

Because the penicillin technique was originally used to enrich for mutants that were mutated on the genome, we decided to replicate this detail of the experiment. We knocked a *pheS* mutant library directly into the genome, replacing the wildtype allele. The resulting mutants were subjected to the penicillin technique negative selection (as described in Figure 1C), and screened for chemical complementation.

It worked. SLiDE strain pheS.GL1 displayed a 10,000 fold decrease in escape frequency (Table 3). After a second round of directed evolution, we generated SLiDE strain pheS.GL2 (Figure 5A-C provides an overview of pheS.GL2 structure and performance). With an escape frequency on the order of 1×10^{-9}, this SLiDE strain was a 100,000 fold improvement over earlier generations (Table 3). This was one of the breakthrough experiments that would mark a shift in our approach to SLiDE strain engineering.

Mutant	Generation	Approximate Escape Frequency
pheS.GLs1	1	1×10^{-1}
pheS.GLf16	10*	1×10^{-4}
pheS.GL1	11	1×10^{-8}
pheS.GL2	12	1×10^{-9}

Table 3. Snapshots of *pheS* Directed Evolution

*Estimated generation

The mutations in pheS.GL2 were very close to the active site (Figure 1B, 3, 5A). Karanicolas suggested that the destabilization of the active site by mutation of buttressing residues (second shell residues) could abolish protein function. Given the proximity of mutations to the active site and the fact that they might be considered second shell residues, it is possible that a mechanism similar to Karanicolas' description could be at play (Figure 3, 5A).

The new methodology that generated SLiDE strain pheS.GL2 provided a benchmark for possible SLiDE strain performance. It established a precedent for the magnitude change in escape frequency that might be possible in a single

round of directed evolution. However, the effectiveness of the new method clearly indicated the limits of SLiDE strain phenotypes.

While the performance of pheS.GL2 was improved over pheS.GL1, the improvements were modest. This suggested that we were reaching a point of diminishing returns for *pheS*-based SLiDE strains. Analysis of escape mutations suggested suppressor mutations were the dominant mechanism of reversion (Figure 3). We interpreted this to mean that SLiDE strain escape frequencies were bounded near the mutation rate of *E. coli*. We concluded that large decreases to escape frequency necessary for biocontainment would require multiple SLiDE alleles. Therefore, we decided to apply our protein engineering methodology to the development of SLiDE strains for other essential genes.

Our first approach to library design was based on the assumption that ligand binding would require a spherical constellation of stabilizing residues to surround the allosteric effector. A visual analogy of this spherical library design approach can be found in the morphology of a dried dandelion. At the center of the dandelion is a solid knob, what we can describe as our ligand. Attached to the dandelion's central knob are hundreds of little seeds. These seeds can be thought of as analogous to the amino acids that might stabilize a ligand in the center of the allosteric pocket. (See the right portion of mutation in figure 6 for an example of spherical/dandelion library design approach.)

While the spherical/dandelion approach successfully generated early examples of ligand dependence, the library design was strictly informed by the examination of tertiary structure. Mutations were often distributed relatively far from one another in DNA sequence. This made the library fabrication more costly in terms of time. The opportunity cost of this design complexity was the exploration of simpler library design approaches.

The structural arrangement of mutations in pheS.GL2 was important in revising our library design methodology for subsequent rounds of SLiDE strain engineering. The two rounds of random mutagenesis resulted in the spatial distribution of mutations changing from spherical to linear. This suggested flexibility in our library design approach. The more linear distribution of mutations suggested not only a diversity of library design strategies in general, but also a much cheaper and easier approach to mutagenesis in specific. The linear library design approach was attractive because it could be generated using a single, relatively short DNA sequence. This simplified design reduced time and cost of library fabrication. Reduced experimental complexity allowed inspection of more

library design methodologies. See Figure 19A for example spherical library and Figure 19B for example linear library.

The creation of new biological activities generally follows a process reflective of natural selection: we generate diversity and then we screen for the desired phenotype. The development of SLiDE strain pheS.GL2 was key in reducing the cost, complexity, and required time for both the generation of diversity and the screening of diversity for desired phenotypes. The increases in efficiency with respect to library design (linear libraries), library fabrication (genome libraries), and selection (penicillin technique), would prove vital in the extension of methodologies developed for pheS.GL2 to other essential genes.

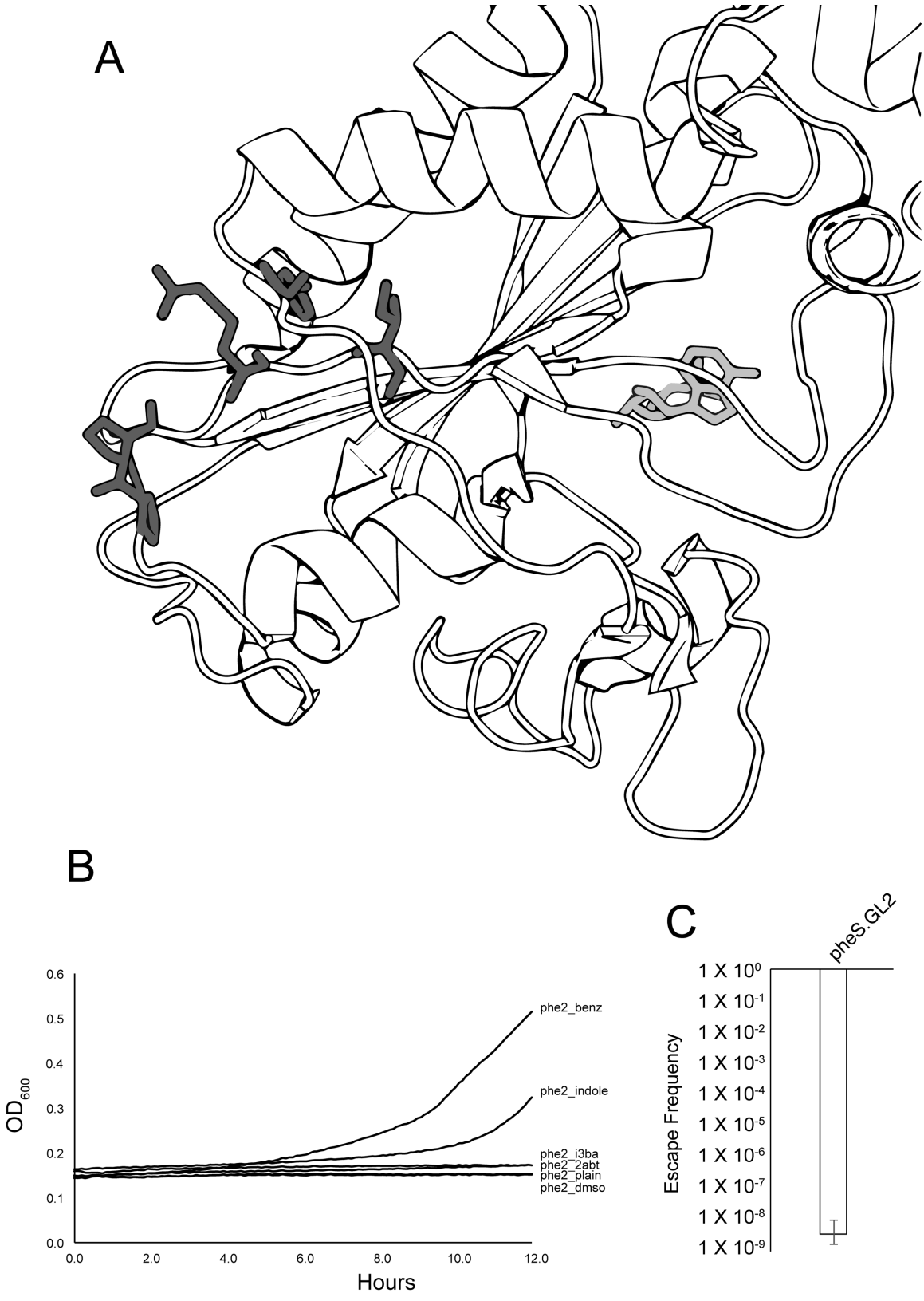

A
B
C
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OD600
0.0
2.0
4.0
6.0
8.0
10.0
12.0
Hours
phe2_benz
phe2_indole
phe2_i3ba
phe2_2abt
phe2_plain
phe2_dmso
pheS.GL2
Escape Frequency
1 X 10^0
1 X 10^-1
1 X 10^-2
1 X 10^-3
1 X 10^-4
1 X 10^-5
1 X 10^-6
1 X 10^-7
1 X 10^-8
1 X 10^-9

Figure 5. SLiDE Strain pheS.GL2

A. The crystal structure of wildtype *pheS* with the residues mutated in SLiDE strain pheS.GL2 highlighted in dark grey. The AMP substrate is highlighted in light gray. The mutations and the substrate are separated by a beta sheet.

B. Growth curve of SLiDE strain pheS.GL2 in various complementing chemicals, DMSO only, or plain media. Concentrations: 1mM benzothiazole (benz), 1mM indole-3-butyric acid (i3ba), 0.5mM 2-aminobenzothiazole (2abt), 0.5 indole, 0.1% DMSO (dmso).

C. The escape frequency of SLiDE strain pheS.GL2.

SLiDE Strain dnaN.GL7

dnaN is unique among the essential genes that we engineered for ligand-dependence because it is not an enzyme. A structural component of the replisome, DnaN protein is a homodimer with two semi circle monomers forming a ring. The ring, also known as “clamp,” is loaded onto double stranded DNA, whereupon other components of the replisome associate and DNA replication can occur(24).

The development of SLiDE strain dnaN.GL7 followed a history similar to pheS.GL2. We made single point mutants in WT *dnaN*, transformed them into *dnaN*^{ts} strains, and then screened the phenotypes for ligand dependence at the restrictive temperature (Figures 13, 15 and Methods Section for details). Several mutants displayed a weak ligand-dependent phenotype. All mutants displaying the styrene-responsive phenotypes were phenylalanine to glycine mutants, just like the *pheS* SLiDE alleles. Because these results were nearly identical to observations made in *pheS*, we concluded that we were on the right path to engineering ligand-dependence into essential genes.

As with *pheS*, *dnaN* underwent many rounds of targeted mutagenesis with two amino acids randomized in each library. The same temperature sensitive strains approach was used for screening. Directed evolution based on small libraries and colony picking did not yield dramatic improvements to *dnaN* mutant performance (Table 4).

Despite several failures with the penicillin technique on *pheS*, the approach was attempted on an error prone PCR library of our best *dnaN* mutant up to that point: dnaN.GLd11 (Table 4 for dnaN.GLd11 and Figure 14 for library fabrication schematic). The resulting library was passed through the penicillin technique dual selection twice (Figures 1C and 15 for dual selection and the temperature sensitive variation). The selection worked so well that a serial dilution of the library (intended to track selection efficiency) showed a more than 1000 fold growth increase when plated on LB agar plates containing a chemical pool.

The original library size was ten million. Screening revealed about 10 different genotypes. There were 1000 times more SLiDE alleles in the library than wt mutants after the selection. Therefore, 1000 (SLiDE/WT) divided by 1×10^{-6} (SLiDE/WT) gives a one billion fold enrichment for SLiDE phenotypes over two rounds of directed evolution.

Where we had been able to screen on the order of 100s to 1000s of mutants with colony picking and replica gridding methodologies, the penicillin technique negative selection allowed us to increase our throughput over 100,000 fold. The penicillin technique became the primary tool for isolating SLiDE alleles, allowing us to probe large libraries.

After generating *dnaN* SLiDE mutants using temperature sensitive and plasmid-based selection techniques, we integrated SLiDE allele dnaN.GL7 into the genome and assessed its performance. SLiDE strain dnaN.GL7 displayed excellent ligand-dependent growth in liquid, complemented by benzothiazole, 2-aminobenzothiazole, and weakly by indole (Figure 2B, 6B). The low escape frequency, on the order of 1 x 10^{-6}, convinced us that large phenotypic changes were possible with this methodology (Figure 6C).

Mutant	Generation	Escape Frequency
dnaN.GLs1	1	1 x 10^{-1} - 1 x 10^{-2}
dnaN.GLd11	5*	1 x 10^{-2} - 1 x 10^{-3}
dnaN.GL7	7	1 x 10^{-6}

Table 4. Snapshots of *dnaN* Directed Evolution

*Estimated generation

The wildtype DnaN protein interacts with its partner proteins through a "hydrophobic cleft." In Figure 6A, we can see two clusters of dark gray residues (positions that were found to be mutated in SLiDE strain dnaN.GL7). The clusters reside on either side of a long, somewhat unfolded polypeptide chain containing a short alpha helical turn. This is the hydrophobic cleft. It is noteworthy that the mutations clustered around the functionally essential hydrophobic cleft. Because the residues reside in the second shell of the hydrophobic cleft Karanicolas' description of buttressing-residue disruption might be relevant.

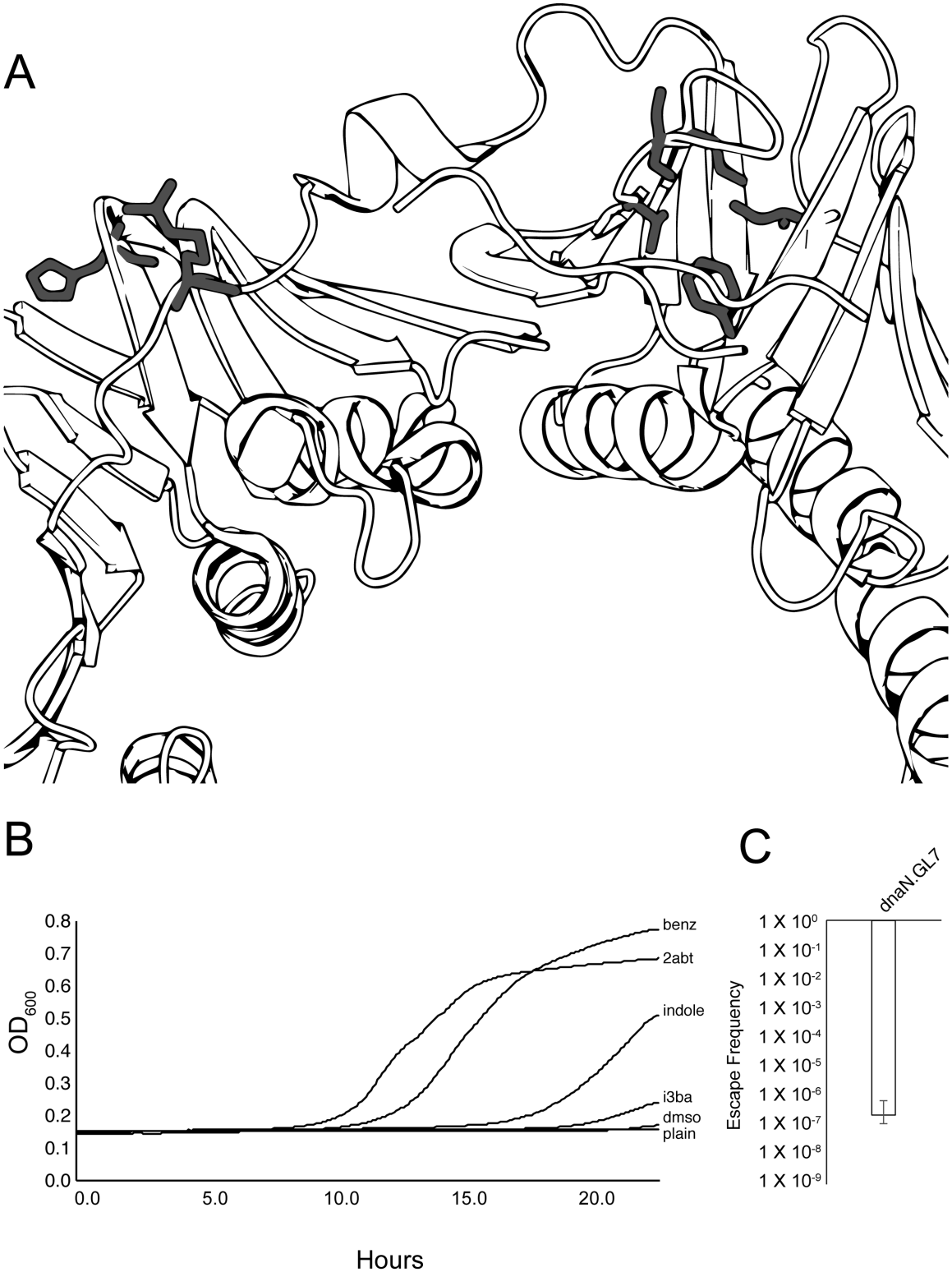
We attempted to characterize the biochemical nature of dnaN.GL7 chemical complementation. To do this, we performed a thermal stability shift assay. Our experiment used a fluorophore, SYPRO orange, to provide a readout of protein thermal stability. SYPRO orange is not fluorescent in an aqueous phase, but fluoresces strongly in a hydrophobic environment. If SYPRO orange is mixed with purified protein (in an aqueous buffer), no fluorescence will be observed. However, as the temperature is increased, the protein will begin to denature.

SYPRO orange will gain access to the hydrophobic core of the protein and can then produce a fluorescent signal. The increase in fluorescent signal is used to determine the extent of a protein's denaturation. The relationship between temperature and denaturation can be used to determine a protein's thermal stability and melting temperature.

We performed the thermal stability assay on both the wildtype DnaN protein and dnaN.GL7. The experiment was performed using a variety of indole concentrations. This allowed us to determine if thermal stability of either wildtype DnaN or dnaN.GL7 were affected by indole. Figure 6D illustrates the thermal stability of WT DnaN (dashed lines) and dnaN.GL7 (solid lines). The readout can be interpreted as the percentage of protein that is denatured (Y-axis) for a given temperature (X-axis). Figure 6E shows the negative derivative of the 6D and the lowest number of the curve defines the melting temperature of a protein. The most obvious feature of Figures 6D and 6E is the large decrease in thermal stability of dnaN.GL7 compared to wildtype DnaN. The more interesting observation, however, is the thermal stability increase of dnaN.GL7 that occurs as the concentration of indole is increased.

The increase in dnaN.GL7 protein stability was caused by the addition of indole (darker lines means more indole), while no effect was observed when only DMSO solvent was added (lightest line, labeled 0 in the legend). The increase in dnaN.GL7 thermal stability displayed a dose-dependent relationship with respect to the concentration of indole added. No such changes were observed for the wildtype protein. This allowed us to conclude that indole increases the thermal stability of dnaN.GL7.

Because of the relationship between protein structure and phenotype, the change in thermal stability must have a structural basis. Therefore, the increase in thermal stability indicates a change in protein conformation. This allows us to conclude that indole mediates a conformational change in dnaN.GL7 that leads to increased thermal stability. In combination with the ligand-dependent growth observed in SLiDE strain dnaN.GL7, the thermal stability shift results provide biochemical evidence for ligand-dependent dnaN.GL7 activity.


A
B
C
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OD600
0.0
5.0
10.0
15.0
20.0
Hours
benz
2abt
indole
i3ba
dmso
plain
dnaN.GL7
Escape Frequency
1 X 10^0
1 X 10^-1
1 X 10^-2
1 X 10^-3
1 X 10^-4
1 X 10^-5
1 X 10^-6
1 X 10^-7
1 X 10^-8
1 X 10^-9

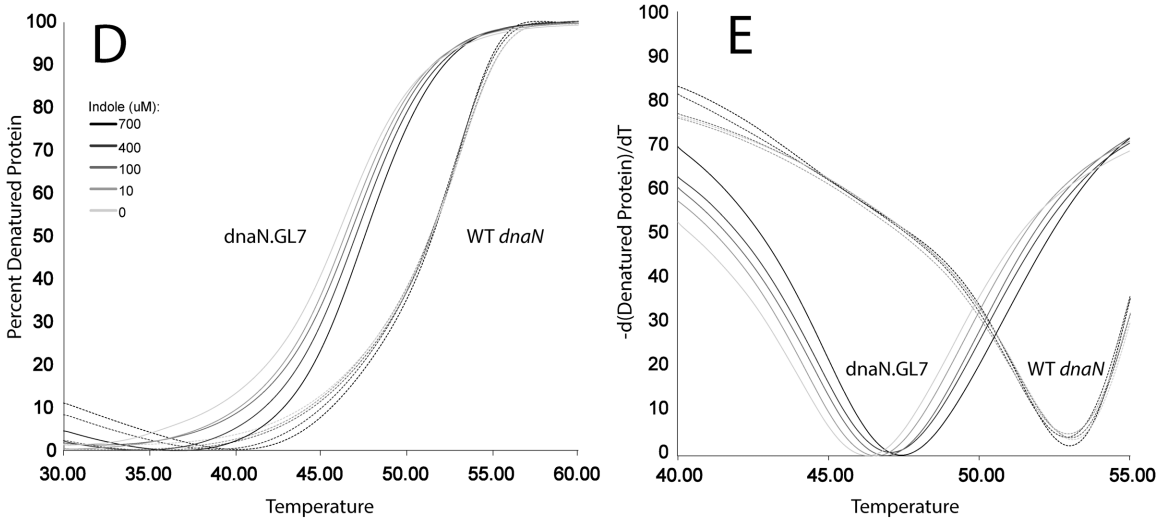


Figure 6. SLiDE Strain dnaN.GL7

A. The crystal structure of wildtype *dnaN* with the residues mutated in SLiDE strain dnaN.GL7 highlighted in dark grey. The hydrophobic cleft, important in replisome assembly, runs between the two patches of mutations.

B. Growth curve of SLiDE strain dnaN.GL7 in various complementing chemicals, DMSO only, or plain media. Concentrations: 1mM benzothiazole (benz), 1mM indole-3-butyric acid (i3ba), 0.5mM 2-aminobenzothiazole (2abt), 0.5 indole, 0.1% DMSO (dmso).

C. The escape frequency of SLiDE strain dnaN.GL7.

D. Thermal shift assay curves. Solid lines show denaturation of dnaN.GL7. Dashed lines show denaturation of WT *dnaN*. Darker lines are higher concentrations of indole, lighter lines are lower concentrations of indole. The lightest color, 0, is DMSO only.

E. Negative derivative of thermal stability. Lowest number indicates maximum rate of protein denaturation and is used to define a protein's melting temperature. Note the increase in thermal stability of dnaN.GL7 with more indole (darker lines to the right). Note the lack of effect in wildtype DnaN.

SLiDE Strains tyrS.GL7 and metG.GL15

SLiDE strains tyrS.GL7 and metG.GL15 were the first mutants generated based on the lessons of library design and improvements of selection and screening methodologies. These mutants established several key features of our SLiDE engineering approach:

1. SLiDE engineering could be done in one step of library fabrication, selection, and screening (in contrast to the iterative approaches used for *pheS* and *dnaN*). See Figure15 for library fabrication overview.

2. SLiDE strains could be generated using the linear library architecture (as opposed to the spherical or dandelion approach discussed in the section on *pheS*). See Figures 7A and 8A for library architecture.

3. SLiDE engineering could be extended beyond *pheS* and *dnaN*.

4. SLiDE engineering could produce phenotypes on the order of weeks (in contrast to the 1-2 years required for *pheS* and *dnaN*).

5. SLiDE engineering worked particularly well for generating indole-like responsive mutants (consistent with ligands observed to mediate complementation in *pheS* and *dnaN*).

6. SLiDE engineering worked particularly well when libraries were targeted near functionally important loci of the targeted enzyme (like the substrate binding pocket of *pheS* or the hydrophobic cleft involved in replisome assembly of *dnaN*). Note the light gray substrates in Figures 7A and 8A.

In sum, these features suggested that our approach might make the engineering of ligand-dependent phenotypes easy. This has implications for the development of biosensors, engineered feedback (as for control of metabolic pathways), and the more general efforts of understanding ligand binding.

Based on the observation that large decreases in escape frequency could be generated (during the development of pheS.GL1, pheS.GL2, and dnaN.GL7), we hypothesized that we could skip the multiple rounds of directed evolution with the right combination of library design, selection, and screening. We hypothesized that generating libraries directly on the genome might yield SLiDE strains without the need for complex plasmid-based screening strategies .

For *tyrS* and *metG* SLiDE engineering, we implemented a different approach to library fabrication (the details of which are summarized in Figure 16 and the methods section). This approach used a fusion of Cas9 cutting and oligo-mediate recombination. This technique created libraries with minimal changes to the genomic context (though the helper plasmids were required).

The streamlined library fabrication approach was enabled by a different library design approach. The observation of the linear distribution of mutations in pheS.GL2 formed the basis for the linear library design approach. Figures 7A and 8A illustrate the three dimensional architecture of the library designs. Note that mutagenesis is targeted in strips. These mutant strips correspond to 21 bp (7 codons) in DNA sequence. The degeneracies and 5' and 3' homology regions could easily be encoded in a 60bp recombineering oligo.

The result of our new approaches to SLiDE engineering was the extension of SLiDE phenotypes to two more essential genes. Figures 7B and 8B show the ligand-dependent growth curves of SLiDE tyrS.GL7 and metG.GL15, respectively. Figures 7C and 8C show the SLiDE strains' escape frequency. The performance of these mutants (tyrS.GL7 in particular) is especially noteworthy, considering that they were developed in under two weeks. This includes fabrication of helper plasmids, QC, library fabrication using the helper plasmids, selection, and screening.

While SLiDE engineering on these genes did not explore the use of different complementing chemicals, it is noteworthy that the same chemicals (Figure 1A) that complemented pheS.GL2 and dnaN.GL7 also worked with tyrS.GL7 and metG.GL15. This raises the question of whether there is something unique about these indole-like molecules for mediating chemical complementation. Perhaps it is something unique to the combination of these molecules and our library design approaches.

Another commonality between these new mutants and the older mutants was the proximity between mutations (dark gray) and the functionally important loci/substrates (light gray) of the proteins (Figure 7A and 8A). While this might not seem so surprising, it is important to note the consistency of this trend in SLiDE strains as a possible design principle for future studies. Karanicolas' concept of "buttressing" and the importance of second shell amino acids must be considered, given their consistent appearance in all four SLiDE strains.

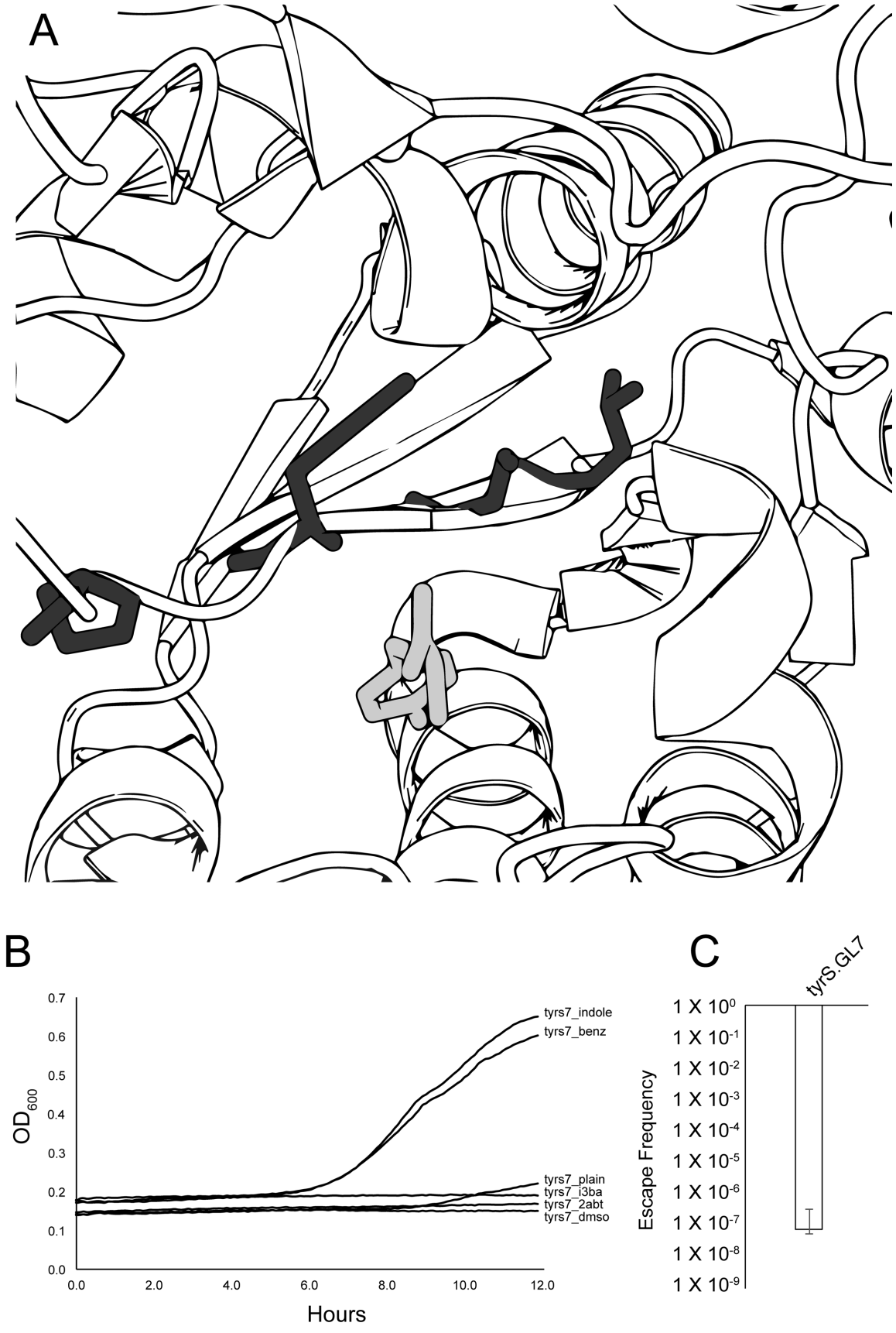

A
B
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OD600
0.0
2.0
4.0
6.0
8.0
10.0
12.0
Hours
tyrs7_indole
tyrs7_benz
tyrs7_plain
tyrs7_i3ba
tyrs7_2abt
tyrs7_dmso
tyrS.GL7
Escape Frequency
1 X 10^0
1 X 10^-1
1 X 10^-2
1 X 10^-3
1 X 10^-4
1 X 10^-5
1 X 10^-6
1 X 10^-7
1 X 10^-8
1 X 10^-9

Figure 7. SLiDE Strain tyrS.GL7

A. The crystal structure of wildtype *tyrS* with the residues mutated in SLiDE strain tyrS.GL7 highlighted in dark grey. The substrate is highlighted in light gray. The mutations and the substrate are separated by a beta sheet.

B. Growth curve of SLiDE strain tyrS.GL7 in various complementing chemicals, DMSO only, or plain media. Concentrations: 1mM benzothiazole (benz), 1mM indole-3-butyric acid (i3ba), 0.5mM 2-aminobenzothiazole (2abt), 0.5 indole, 0.1% DMSO (dmso).

C. The escape frequency of SLiDE strain tyrS.GL7.

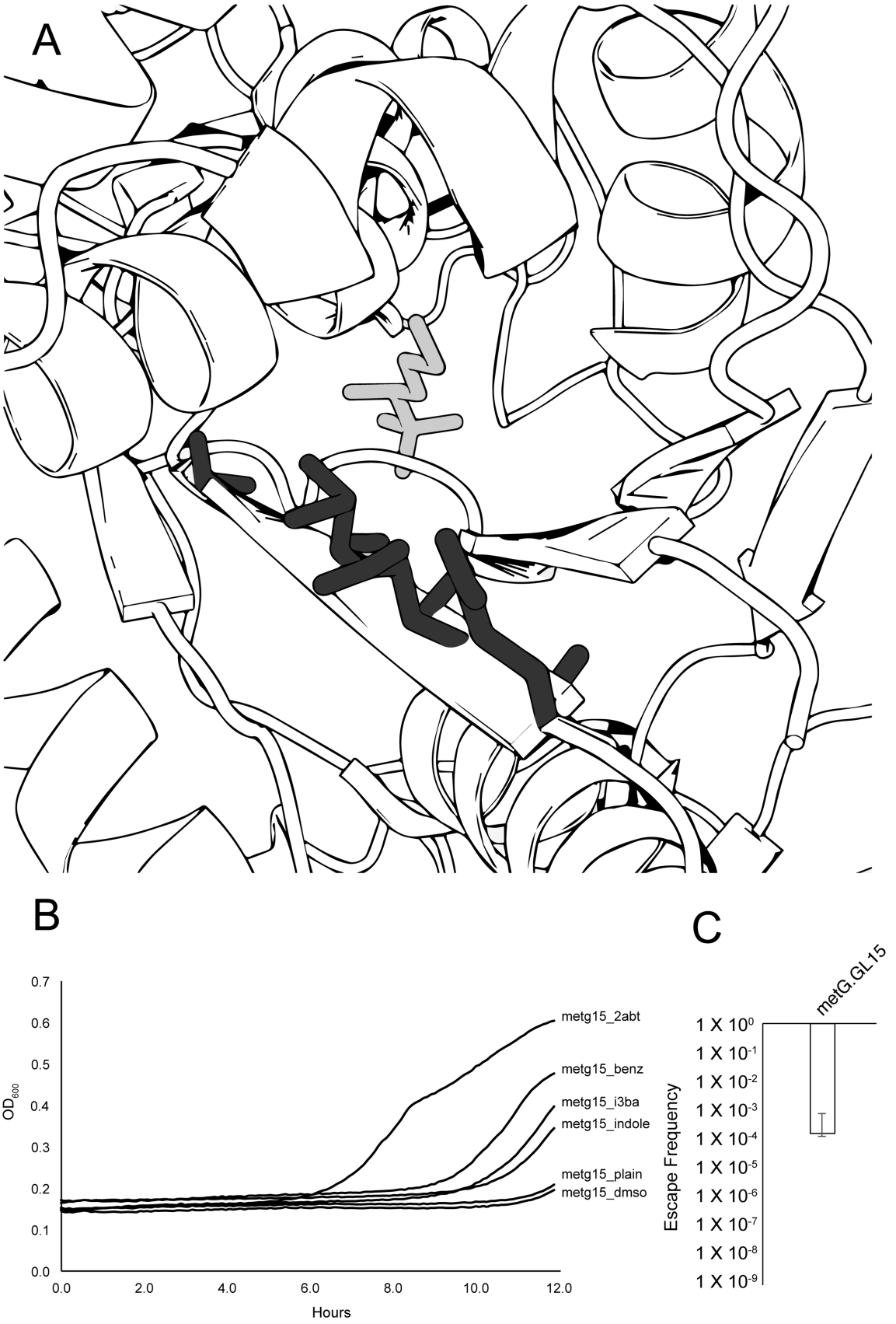

A
B
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OD600
0.0
2.0
4.0
6.0
8.0
10.0
12.0
Hours
metg15_2abt
metg15_benz
metg15_i3ba
metg15_indole
metg15_plain
metg15_dmso
metG.GL15
Escape Frequency
1 X 10^0
1 X 10^-1
1 X 10^-2
1 X 10^-3
1 X 10^-4
1 X 10^-5
1 X 10^-6
1 X 10^-7
1 X 10^-8
1 X 10^-9

Figure 8. SLiDE Strain metG.GL15

A. The crystal structure of wildtype *metG* with the residues mutated in SLiDE strain metG.GL15 highlighted in dark grey. The substrate is highlighted in light gray. The mutant residues occupy the hydrophobic core immediately adjacent to the substrate-binding pocket.

B. Growth curve of SLiDE strain metG.GL15 in various complementing chemicals, DMSO only, or plain media. Concentrations: 1mM benzothiazole (benz), 1mM indole-3-butyric acid (i3ba), 0.5mM 2-aminobenzothiazole (2abt), 0.5 indole, 0.1% DMSO (dmso).

C. The escape frequency of SLiDE strain metG.GL15.

SLiDE Strains adk.GL1 and orn.GL1

While the utility of our SLiDE engineering approach seemed to have been validated by tyrS.GL7 and metG.GL15, we aimed to explore more efficient methods for generating SLiDE strains. Additionally, we wanted to ensure that the ligand-dependent phenotypes were not a peculiarity of tRNA ligases (on which three out of three of our SLiDE strains were based: pheS, tyrS, metG). Towards this end, we generated SLiDE libraries on *adk* and *orn*. Both libraries yielded SLiDE strains.

The essential gene *adk*, also known as adenylate cyclase, catalyzies the cyclization of ATP to cyclic AMP (cAMP)(24). The essential gene *orn,* also known as oligoribonuclease, is *E. coli's* only ribonuclease responsible for the degradation of small (2-5 bp) ssDNA and RNA fragments as part of nucleotide recycling(24).

SLiDE strains adk.GL1 and orn.GL1 were the result of a library design strategy constrained by the practicality of library fabrication. See Figure 16 for an overview of the library fabrication methodology and the methods section for a detailed description of the library fabrication approach. The objective of this SLiDE strain engineering approach was to create desired phenotypes using a single PCR fragment and lambda-red integration, followed by our selection and screening methods. We hypothesized that if we were able to integrate a selectable marker immediately upstream up an essential gene, we might be able to mutagenize a portion of the 5' of the essential gene closest to the selectable marker. Therefore, essential genes were inspected for the presence of a 5' DNA sequence that coded for hydrophobic beta strand secondary structure, passing through tertiary structure of high hydrophobicity.

As can be seen in Figure 9A, the mutations in adk.GL1 (in dark gray) all reside on the N-terminus of the protein, with the N-terminal tail clearly visible at the mid-right side of the illustration. In Figure 10A, the mutations (dark gray) also all reside on the N-terminal portion of the peptide sequence. The targeted mutations include amino acid residues 7, 9, 11, and 13 (from right to left in the image). In both adk.GL1 and orn.GL1, the N-terminal peptide chain targeted for mutagenesis has a beta strand secondary structure that immediately enters an area of high hydrophobicity.

The proximity of mutations in adk.GL1 with respect to the enzyme substrates (light gray) are quite obvious. The proximity of mutations and substrates is

consistent with all of the SLiDE mutations seen up to this point. In orn.GL1, mutation 13 (Figure 10A dark gray, top left) overlaps residues 12 and 14 (light gray), which are involved in the metal ion binding necessary for catalytic activity(24). The mutations (dark gray) were introduced on the top face of the beta sheet (with respect to the image), while the metal ion binding residues occupy the bottom face of the beta sheet. The active site is indicated by the catalytic arginine(24) (light gray residue – lowest and furthest to the right of indicated residues).

In addition to DNA sequence-based library design constraints, we tested a new library design approach on adk.GL1 by randomizing residues on both sides of the beta sheet (as opposed to only one side). Because we mutated residues on both sides, we were able to mutate four codons in a row, allowing for a more "compact" library. The "compactness" of the adk.GL1 library was intended to mediate a more spatially focused set of diversity of perturbations to the protein's tertiary structure. This is in contrast to our previous libraries, whose mutations were more distributed. The success of both of these library designs strategies suggest that the relative "compactness" of a potential library design may be a worthwhile variable to consider in future protein engineering endeavors.

The performance of adk.GL1 was on par with SLiDE strain metG.GL15 in terms of escape frequency (Figure 10C versus 8C). However, adk.GL1 appeared to be less promiscuous (Figure 10B versus 8B), suggesting that it might make a superior starting point for generating SLiDE strains with more specific chemical complementation.

SLiDE strain orn.GL1, on the other hand, did not display the same high performance phenotype of other SLiDE strains. The high background growth of orn.GL1 in the absence of complementing ligand was the major weakness of this SLiDE strain (Figure 10B, 10D). The change in growth mediated by benzothiazole was calculated to be 2.5 fold greater using densitometry measurements of the SLiDE strains grown on solid media (Figure 10C, 10D). This is in contrast to higher performance SLiDE strains that displayed almost no growth in the absence of chemicals (Figure 1E). We interpret the small differences in liquid growth curves to be a reflection of high background growth in the absence of complementing chemicals (Figure 10B). Despite orn.GL1's weak performance, we consider it an important piece of evidence as to the extensibility of our SLiDE engineering approach to diverse types of protein structures and enzymatic activities.

One of the advantages of this method for generating libraries is that fewer accessory plasmids and machinery are required. CRISPR-Cas9, while a very powerful tool for generating SLiDE libraries deep into essential gene ORFs, adds a somewhat unwelcome layer of logistical complexity and cost to the experiment (fabrication of help plasmids). Because pSIM5 is the only helper plasmid involved and a single PCR is the only DNA fabrication step required, the experimental simplicity of this approach may be difficult to surpass.

The sum of our library design and fabrication optimizations allowed us to go from a wildtype *E. coli* MC1061 to two SLiDE strains based on two different essential genes in five days and required less than $100 USD. The only steps were a PCR to generate an integration fragment, the competent cell prep for the integration, the positive selection, the negative selection, and screening. This approach sets a benchmark for cost, time, and complexity of future ligand-dependence engineering studies.

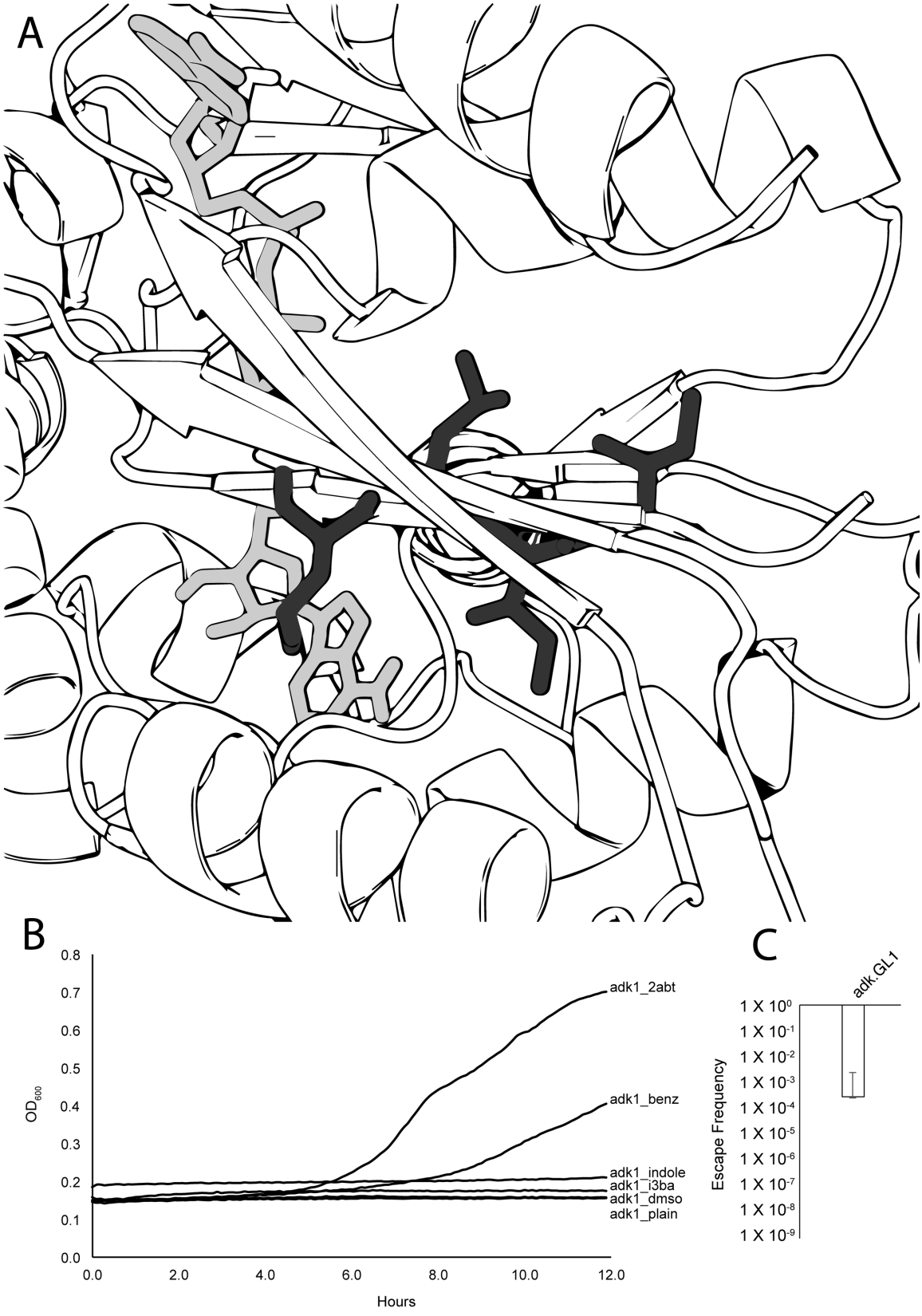

A
B
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OD600
0.0
2.0
4.0
6.0
8.0
10.0
12.0
Hours
adk1_2abt
adk1_benz
adk1_indole
adk1_i3ba
adk1_dmso
adk1_plain
C
adk.GL1
Escape Frequency
1 X 10^0
1 X 10^-1
1 X 10^-2
1 X 10^-3
1 X 10^-4
1 X 10^-5
1 X 10^-6
1 X 10^-7
1 X 10^-8
1 X 10^-9

Figure 9. SLiDE Strain adk.GL1

A. The crystal structure of wildtype *adk* with the residues mutated in SLiDE strain adk.GL1 highlighted in dark grey. The substrates are highlighted in light gray.

B. Growth curve of SLiDE strain adk.GL1 in various complementing chemicals, DMSO only, or plain media. Concentrations: 1mM benzothiazole (benz), 1mM indole-3-butyric acid (i3ba), 0.5mM 2-aminobenzothiazole (2abt), 0.5 indole, 0.1% DMSO (dmso).

C. The escape frequency of SLiDE strain adk.GL1.

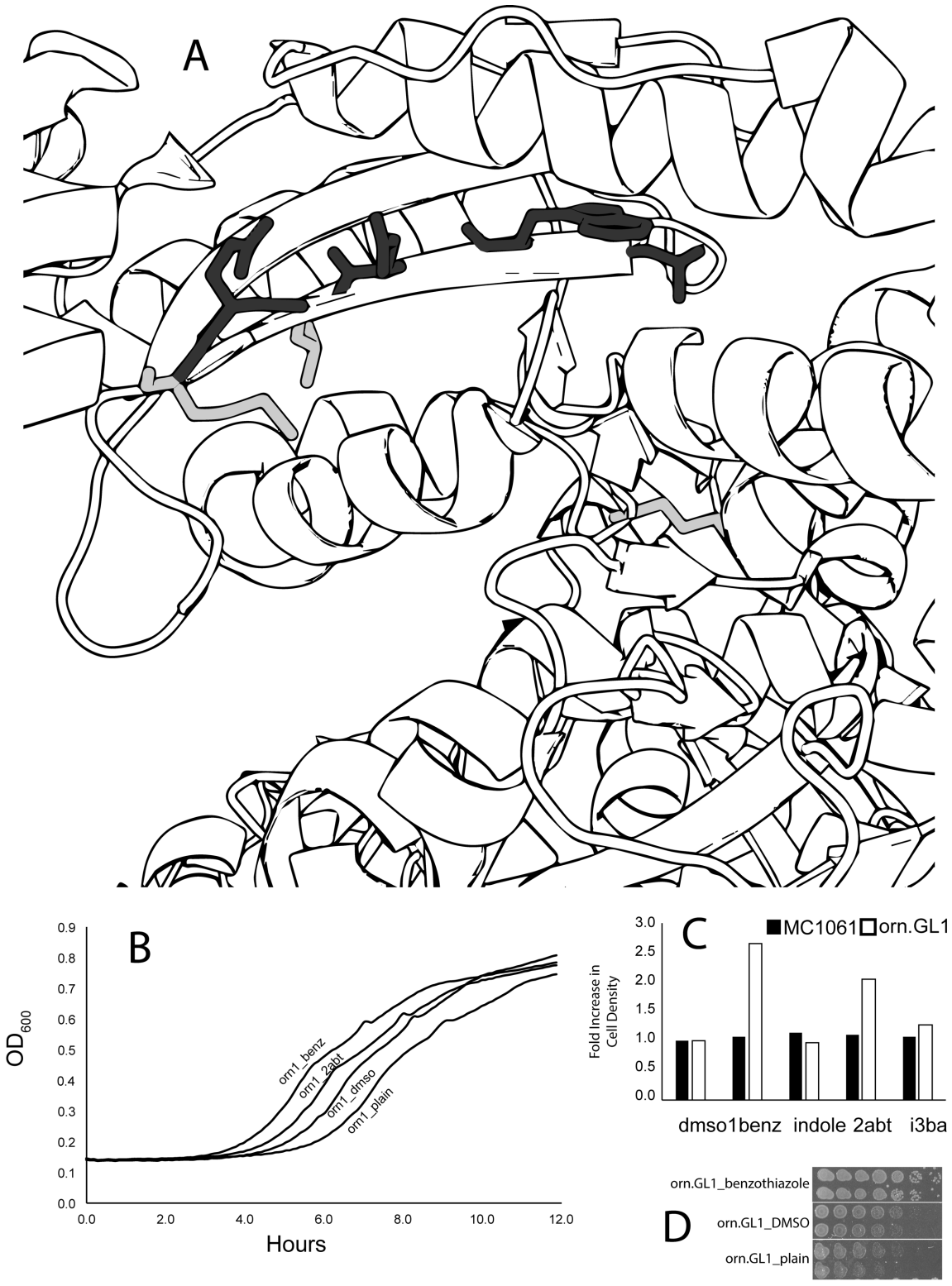

A
B
OD600
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
2.0
4.0
6.0
8.0
10.0
12.0
Hours
orn1_benz
orn1_2abt
orn1_dmso
orn1_plain
C
MC1061
orn.GL1
Fold Increase in Cell Density
3.0
2.5
2.0
1.5
1.0
0.5
0.0
dmso1benz
indole
2abt
i3ba
D
orn.GL1_benzothiazole
orn.GL1_DMSO
orn.GL1_plain

Figure 10. SLiDE Strain orn.GL1

A. The crystal structure of wildtype *orn* with the residues mutated in SLiDE strain orn.GL1 highlighted in dark grey. From right to left in the image, mutated residues occupy positions 7, 9, 11, 13. The pair of metal ion binding residues (12 and 14 from right to left) are on the left of the image, below the mutated residues and are highlighted in light gray. The active site arginine is the lowest and right most residue highlighted in light gray.

B. Growth curve of SLiDE strain orn.GL1 in various complementing chemicals, DMSO only, or plain media. Concentrations: 1mM benzothiazole (benz), 0.5mM 2-aminobenzothiazole (2abt), 0.1% DMSO (dmso). Note that DMSO mediates slight complementation.

C. Relative growth density of orn.GL1 on LB agar plates of various chemical conditions with respect to LB agar plates containing DMSO. Numbers generated from densitometry analysis of serial dilutions.

D. Example raw data used for densitometry analysis in C. Note that DMSO mediates slight complementation with respect to plain LB agar plates.

Combination of SLiDE Mutations to Facilitate Intrinsic Biological Containment With Industrially Relevant Strain BL21(DE3)

Once we felt we had sufficiently demonstrated our ability to generate SLiDE strains, we switched of to demonstrating their applications. The utility of SLiDE alleles for biological containment of a genetically engineered organism was illustrated by combining two and then three sets of SLiDE alleles into BL21(DE3), a strain commonly used in industrial settings (Figure 11). We hypothesized that multiple SLiDE mutations would reduce escapes due to suppressor mutations. It has been shown that multiple, distributed synthetic auxotrophies can reduce escapes due to horizontal gene transfer(18, 19).

Several of our SLiDE strains were fully complemented by 500 μM indole (Figure 2B), which can be generated by the tryptophanase activity of TnaA(28). Although our SLiDE strains did not appear to be complemented by WT *tnaA*, we preemptively eliminated this possible mode of epistatic escape by removing *tnaA*.

Combining tyrS.GL7 (escape frequency 7×10^{-8}) and metG.GL15 (escape frequency 3×10^{-4}), we generated a double SLiDE strain with an escape frequency of 5×10^{-10} (Figure 12A), exceeding the biosafety threshold of one escape mutant per 1×10^{8} cells(19). The escape frequency steadily increased over the first four days, increasing nearly 1000 fold to 4×10^{-7}. From day four to day ten, the escape frequency increased ten fold to 1×10^{-6}. The double SLiDE strain displayed dose-dependent growth in liquid culture in microplate-based growth curve experiments (Figure 12B). The performance of the double SLiDE strain should be considered in light of the development time. Both SLiDE alleles used above were generated in a single round of mutagenesis, selection, and screening over two weeks.

In order to further reduce escape frequency, pheS.GL2 was transferred by P1 transduction into the double SLiDE strain to generate a triple SLiDE strain. The triple SLiDE strain showed mixed colony sizes and we consistently characterized small colonies. The triple SLiDE strain escape frequency dropped below the limit of detection of 3×10^{-11} on days one and two. As with the double mutant, the escape frequency increased over the duration of the experiment, stabilizing at 2×10^{-7} after ten days. The triple SLiDE strain displayed dose-dependent growth in liquid culture, but required a higher concentration of benzothiazole for full chemical complementation (Figure 12D).

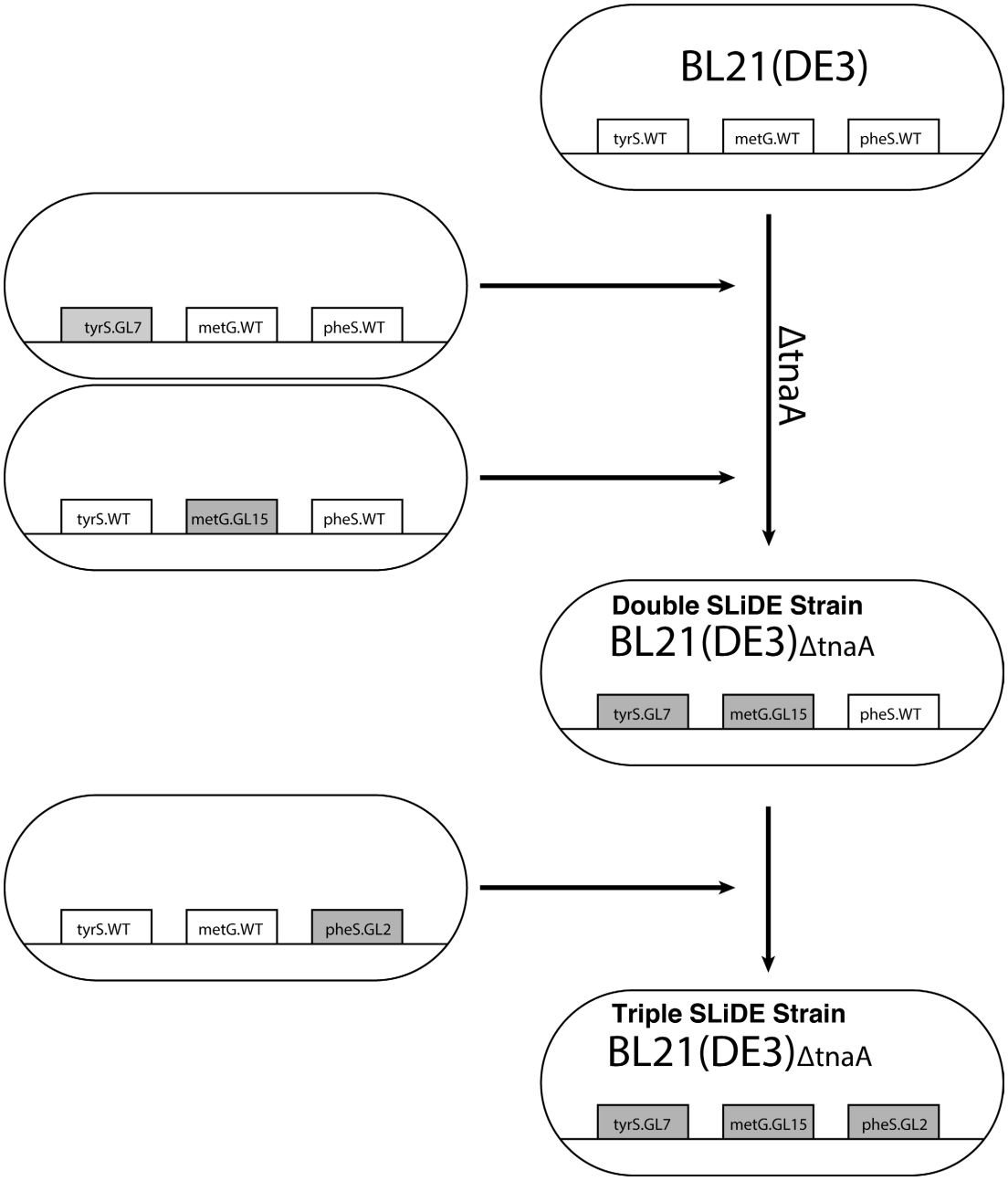
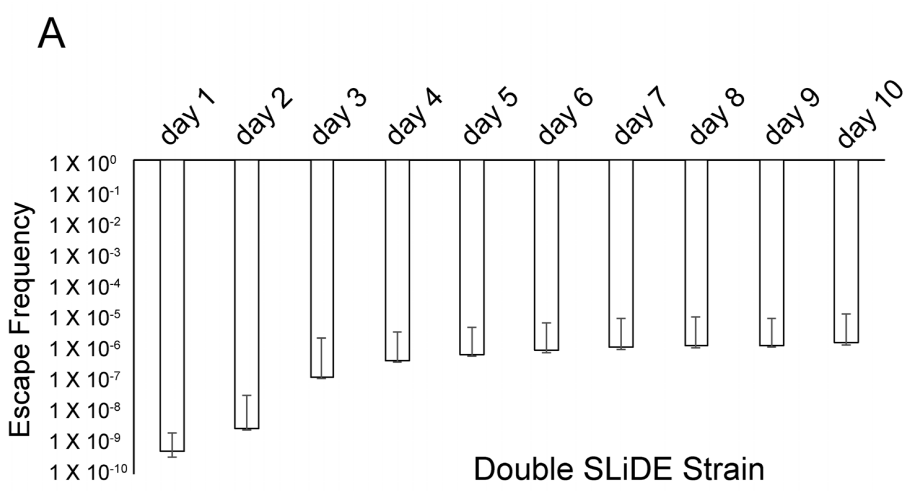
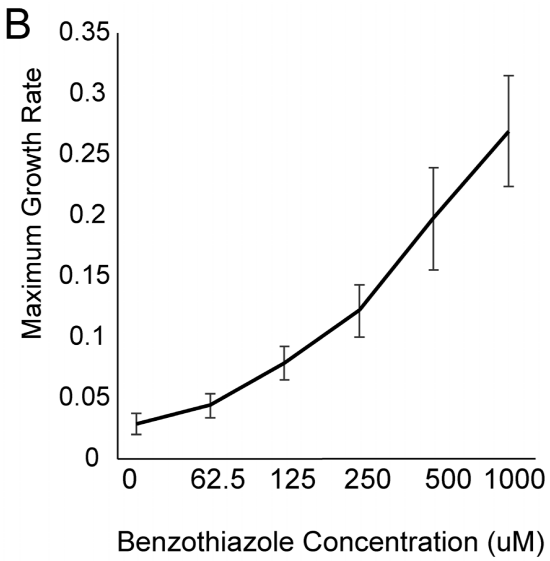
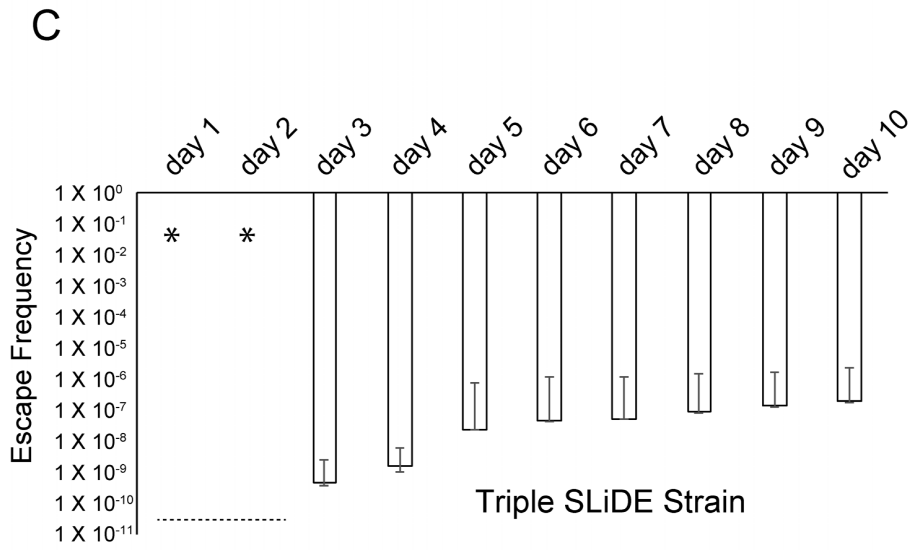

BL21(DE3)
tyrS.WT
metG.WT
pheS.WT
tyrS.GL7
metG.WT
pheS.WT
tyrS.WT
metG.GL15
pheS.WT
ΔtnaA
Double SLiDE Strain
BL21(DE3)ΔtnaA
tyrS.GL7
metG.GL15
pheS.WT
tyrS.WT
metG.WT
pheS.GL2
Triple SLiDE Strain
BL21(DE3)ΔtnaA
tyrS.GL7
metG.GL15
pheS.GL2

Figure 11. Double and Triple SLiDE Strain Fabrication

From top to bottom, each SLiDE module was added to BL21(DE3) serially using P1-transduction or CRISPR-Cas9-assisted Recombineering. The resulting Double and Triple SLiDE strains were used in subsequent biosafety experiments.


A
day 1
day 2
day 3
day 4
day 5
day 6
day 7
day 8
day 9
day 10
1 X 10^0
1 X 10^-1
1 X 10^-2
1 X 10^-3
1 X 10^-4
1 X 10^-5
1 X 10^-6
1 X 10^-7
1 X 10^-8
1 X 10^-9
1 X 10^-10
Escape Frequency
Double SLiDE Strain


B
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Maximum Growth Rate
0
62.5
125
250
500
1000
Benzothiazole Concentration (uM)


C
day 1
day 2
day 3
day 4
day 5
day 6
day 7
day 8
day 9
day 10
1 X 10^0
1 X 10^-1
1 X 10^-2
1 X 10^-3
1 X 10^-4
1 X 10^-5
1 X 10^-6
1 X 10^-7
1 X 10^-8
1 X 10^-9
1 X 10^-10
1 X 10^-11
*
*
Escape Frequency
Triple SLiDE Strain

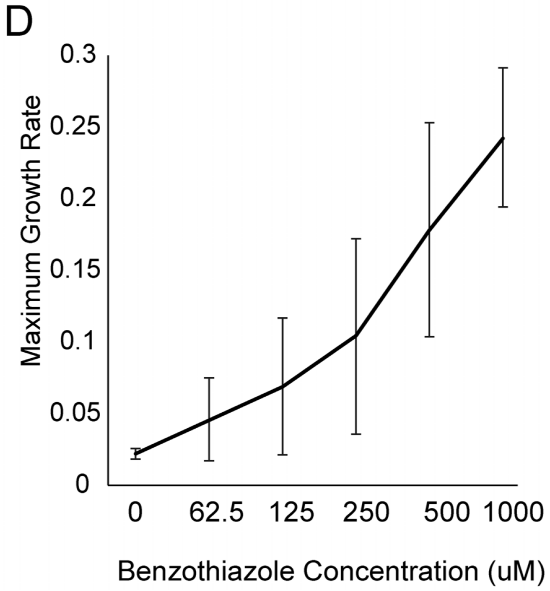

D
0.3
0.25
0.2
0.15
0.1
0.05
0
Maximum Growth Rate
0
62.5
125
250
500
1000
Benzothiazole Concentration (uM)

Figure 12. Double and Triple SLiDE Strain Performance

A. Escape frequencies of tyrS.GL7 / metG.GL15 double SLiDE strain over ten days.

B. Growth rate of tyrS.GL7 / metG.GL15 double SLiDE strain in response to benzothiazole titration.

C. Escape frequencies of tyrS.GL7 / metG.GL15 / pheS.GL2 triple SLiDE strain over ten days. The limit of detection was calculated as the sum of CFU of five biological replicates. Once colonies appeared, the mean and standard deviation were calculated from the five biological replicates.

D. Growth rate of tyrS.GL7 / metG.GL15 / pheS.GL2 triple SLiDE strain in response to benzothiazole titration.

Discussion

Out of the ~300 essential genes in *E. coli*(20), we targeted nine essential genes for mutagenesis, using eleven libraries. Of these, six essential genes yielded SLiDE strains of which we characterize one mutant each (Figure 2B, 5, 6, 7, 8, 9, 10). We conclude that ligand-dependent phenotypes are not rare and can be readily generated using classical microbial genetics techniques.

The use of new technologies can be limited by high cost or high complexity(29). After developing a basic methodology with our first SLiDE strains (dnaN.GL7 and pheS.GL2), we focused on optimizing the protein engineering process. A combination of techniques from the past and present enabled the low cost and rapid development of SLiDE strains. The penicillin technique, for instance, increased the capacity of our directed evolution efforts. Our next mutants (tyrS.GL7, metG.GL15) were generated in two weeks with a single round of mutagenesis, selection, and screening. To further reduce cost and time we simplified library generation and our final SLiDE strains (adk.GL1 and orn.GL1) were developed in five days and for less than $100 in materials. This indicates that synthetic auxotrophs can be developed simply, cheaply, and with a variety of library design approaches.

We demonstrated the biosafety application of SLiDE strains in an industrially relevant strain BL21(DE3), commonly used in large-scale fermentations(22). Biosafety strains were created using off-the-shelf SLiDE alleles with no downstream optimization. Thus, industrially deployable biosafety strategies can be developed with SLiDE strains.

Futures studies may seek to improve performance of SLiDE strains for biosafety by increasing sensitivity, specificity, or the diversity of complementing molecules. The extension of SLiDE strains to other organisms may broaden their industrial utility. Exploring the utility of SLiDE strains for sensor-selector applications could benefit enzyme engineering efforts. Because we demonstrate that SLiDE strains as a rapid and low-cost approach to biological containment we hope our approach reduce the barriers to wider application of synthetic biology technologies.

Methods

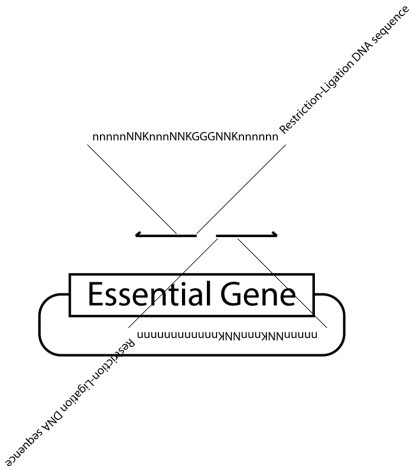
Materials and Equipment

LB agar (Difco) and 2-YT broth (VWR) were used to culture cells at 37 ºC, unless otherwise indicated. Chemicals for ligand screening were purchased from Sigma Aldrich, Spectrum Chemicals, Fisher, and Santa Cruz and were dissolved in DMSO as 1 M stocks and used unpurified. With the exception of 2-aminobenzothiazole and indole, which were used as 500 µM stocks, all stocks were used at 1 mM final concentration. For the negative selection, penicillin (purchased from Spectrum Chemicals) was used at 100 µg/ml final concentration.

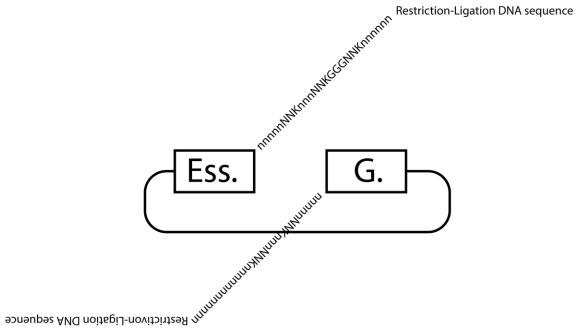
Plasmid Library fabrication

Plasmid libraries on *pheS* and *dnaN* (template plasmid sequence agg119 and agg120) were fabricated using targeted mutagenesis with mutagenic oligos that introduced degeneracies into desired loci. The essential gene template was harbored in a pUC/Spectinomycin plasmid (agg71) that also contained *cre* under pBAD control, although this was not involved in the experiment. The libraries were transformed into temperature sensitive strains (acquired from Yale's Coli Genetic Stock Center) $dnaN^{ts}$ (CGSC# 6844) and $pheS^{ts}$ (CGSC# 4913) at the permissive temperature 30ºC. Screens and selections were performed at the restrictive temperature, 42ºC, in order to uncover the phenotype of the mutant allele.

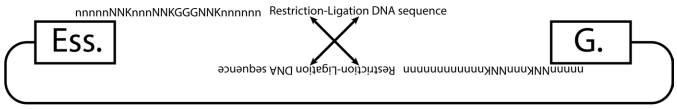
A


Restriction-Ligation DNA sequence
nnnnnNNKnnnNNKGGGNNKnnnnnnn
Essential Gene
nnnnnNNKnnnNNKnnnnnnnnnnn
Restriction-Ligation DNA sequence

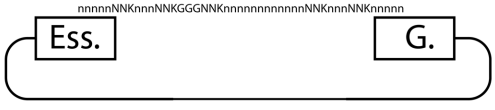
B


Restriction-Ligation DNA sequence
nnnnnNNKnnnNNKGGGNNKnnnnnnn
Ess.
G.
nnnnnNNKnnnNNKnnnnnnnnnnn
Restriction-Ligation DNA sequence

C


nnnnnNNKnnnNNKGGGNNKnnnnnnn
Restriction-Ligation DNA sequence
Ess.
G.
nnnnnNNKnnnNNKnnnnnnnnnnn
Restriction-Ligation DNA sequence

D


nnnnnNNKnnnNNKGGGNNKnnnnnnnnnnnnNNKnnnNNKnnnnn
Ess.
G.

E

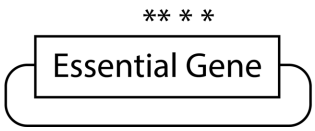

* * * *
Essential Gene

Figure 13. Targeted Mutagenesis Plasmid Libraries

A. A plasmid-borne copy of the targeted essential gene serves as template for an EIPCR (inverse PCR) amplification step with degenerate codons encoded in non-homologous regions of the oligos.

B. The resulting linear PCR fragment from the EIPCR amplification step.

C. The linear DNA is re-circularized using the desired restriction-ligation approach. In our case, we used a golden-gate style re-circularization.

D. The re-circularized plasmid, illustrating incorporated mutations.

E. Cartoon representation of the re-circularized plasmid. See Figure 15 for temperature sensitive screening and selection strategy.

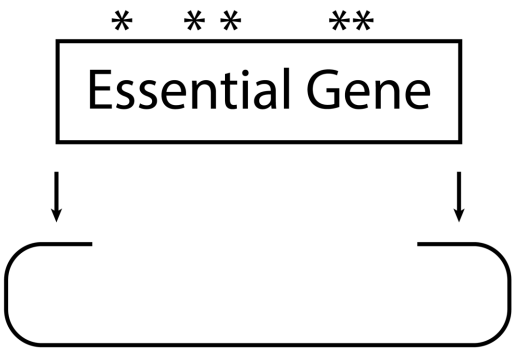
A


Essential Gene

B


* * * **
Essential Gene

C


* * * **
Essential Gene

D


* * * **
Essential Gene

Figure 14. Error Prone Plasmid Mutagenesis

A. Essential gene (or SLiDE allele) template with standard subcloning primers.

B. Resulting mutagenized DNA fragment from error prone PCR.

C. Restriction-ligation subcloning of mutagenized fragment into plasmid backbone.

D. Randomly mutagenized essential gene library for use in temperature sensitive screening and selection strategy (Figure 15).

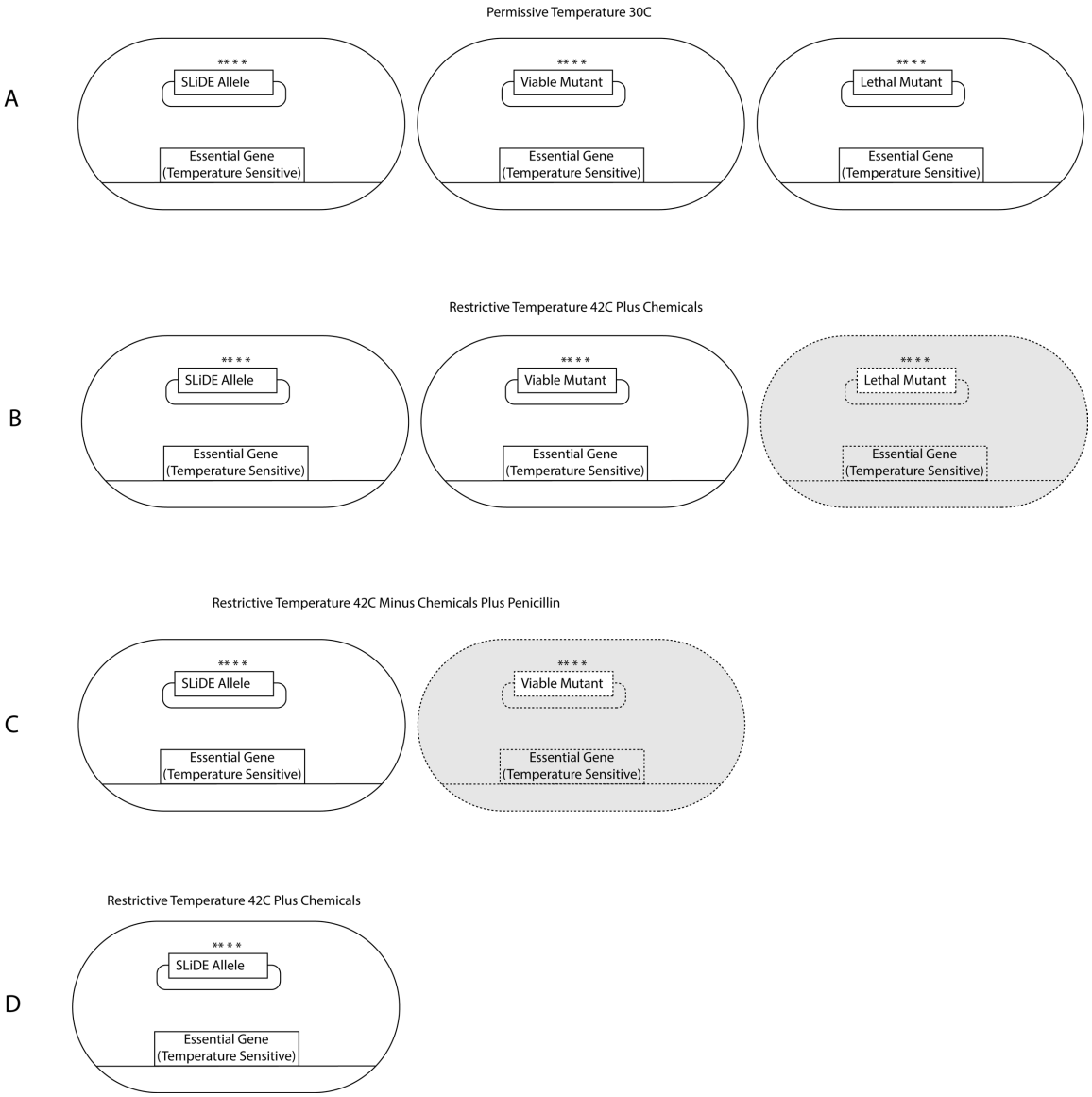

A
Permissive Temperature 30C
SLiDE Allele
Essential Gene (Temperature Sensitive)
Viable Mutant
Essential Gene (Temperature Sensitive)
Lethal Mutant
Essential Gene (Temperature Sensitive)
B
Restrictive Temperature 42C Plus Chemicals
SLiDE Allele
Essential Gene (Temperature Sensitive)
Viable Mutant
Essential Gene (Temperature Sensitive)
Lethal Mutant
Essential Gene (Temperature Sensitive)
C
Restrictive Temperature 42C Minus Chemicals Plus Penicillin
SLiDE Allele
Essential Gene (Temperature Sensitive)
Viable Mutant
Essential Gene (Temperature Sensitive)
D
Restrictive Temperature 42C Plus Chemicals
SLiDE Allele
Essential Gene (Temperature Sensitive)

Figure 15. Temperature Sensitive Screening and Selection Strategy

A. A plasmid library (from Figure 13 or 14) is transformed into the correct temperature sensitive mutant. Cells are grown at the permissive temperature. All library members survive.

B. Cells are spread onto LB agar plates containing the desired chemical pool and grown at the restrictive temperature. At the restrictive temperature, the genomic copy of the essential gene is no longer functional and viability is determined by the plasmid-borne library member. SLiDE strains (chemically complemented) and viable mutants are able to complement the temperature sensitive allele. The SLiDE allele is involved in two levels of complementation – chemical complementation and genetic complementation. Lethal mutants are unable to complement the genomic allele and die (gray).

C. The library is washed of complementing chemicals and then grown without chemicals but with penicillin. Viable mutants start to divide and are killed (gray) by penicillin (Figure 1C). SLiDE alleles don’t grow and don’t die.

D. The library is washed of penicillin and then grown with the addition of complementing chemicals at the restrictive temperature in order to enrich for SLiDE alleles.

Genome Library Fabrication with CRISPR-Cas9

Genome libraries designed by randomizing strips of neighboring amino acids were integrated directly into the genome by recombineering(30). We used a degenerate library oligo to introduce mutations into the genome. We used a CRISPR-Cas9-mediated(31) double strand break targeted to wildtype DNA sequences to enrich for organisms mutagenized at the desired locus. Upon transformation the resulting library of strains were subjected to selections and screens.

Libraries were encoded on 60-bp recombineering oligos (library oligos). Library oligos (478_GLU45.1 for *metG* libraries and 475_LEU36.1 for *tyrS* libraries) contained a 5' homology region of 21-bp, a 21-bp window for degeneracies, and an 18-bp 3' homology arm. Library-containing-recombineering-oligos were accompanied by a helper plasmid (bgg472), encoding an sgRNA targeted to the wildtype sequence to which the degenerate library oligo was homologous. The degenerate library repair oligo and its partner sgRNA plasmid were co-transformed into freshly prepared, electrocompetent MC1061 cells harboring pSIM5(32) encoding the lambda red genes and DS-SPcas encoding Cas9 (33). Cells were recovered for one hour in 2YT and then spread on LB agar plates containing the ligand or pool of interest and antibiotics selecting for DS-SPcas (spectinomycin 50µg/ml final concentration) and the sgRNA plasmid (kanamycin 50µg/ml final concentration). During this incubation, the addition of complementing chemical was found to be optional.

Kanamycin resistant cells (expressing the sgRNA intended to eliminate WT, non-recombinant essential gene background), were enriched for cells in which the targeted locus had been mutated by the library oligo; these cells were treated as the SLiDE strain library. Libraries were grown on solid media instead of liquid media to ensure that slowly growing mutant strains (with desirable ligand-dependent phenotypes) would not be outcompeted by faster growing mutants, lacking ligand-dependence. The resulting library was then subjected to selection and screening.

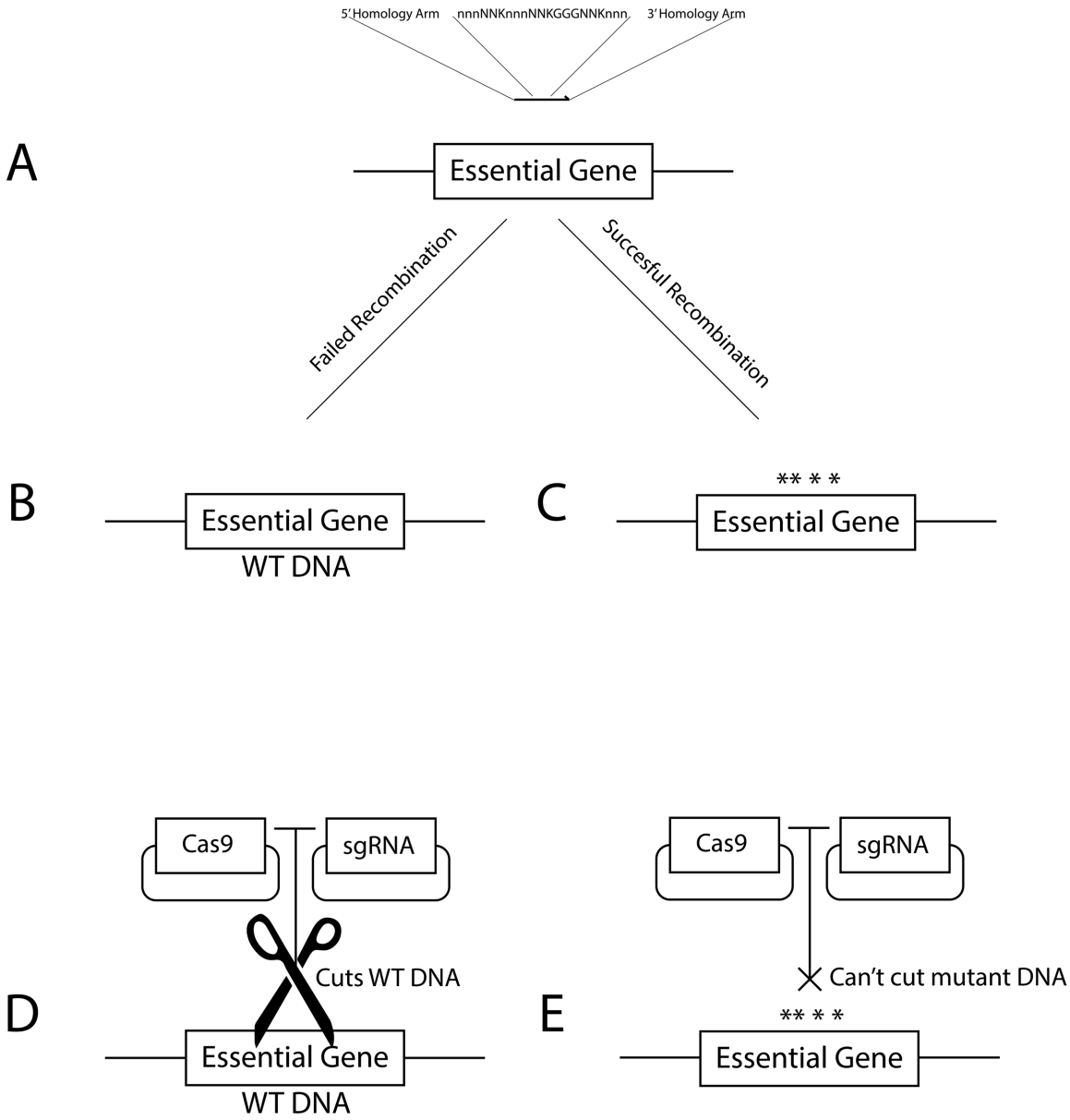

5' Homology Arm
nnnNNKnnnNNKGGGNNKnnn
3' Homology Arm
A
Essential Gene
Failed Recombination
Succesful Recombination
B
Essential Gene
WT DNA
C
** * *
Essential Gene
D
Cas9
sgRNA
Cuts WT DNA
Essential Gene
WT DNA
E
Cas9
sgRNA
Can't cut mutant DNA
** * *
Essential Gene

Figure 16. CRISPR-Cas9 Recombineering Libraries

A. A library oligo containing degeneracies targeted for an essential gene is transformed into cells harboring lambda-red and Cas9 machinery (not shown).

B. If library integration is not successful, the essential gene remains wildtype and susceptible to cleavage by an sgRNA targeted to wildtype sequence (not shown).

C. If library integration is successful, then the essential gene's sequence is altered and will not be recognized by an sgRNA targeted to wildtype sequence (not shown)

D. The un-mutated essential gene is cut by Cas9 activity, killing cells the desired degeneracies.

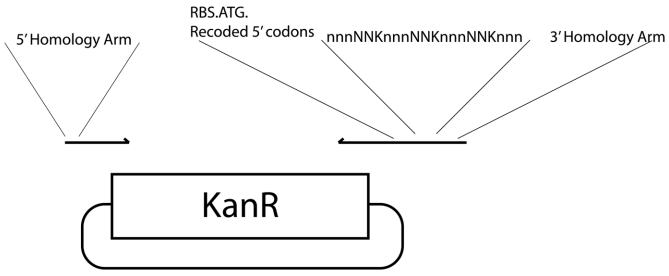
E. Essential genes that were mutated can no longer be recognized by Cas9. The essential gene is not cut and the cell can undergo selection and screening.

Genome Library Fabrication with Lambda-Red Recombination

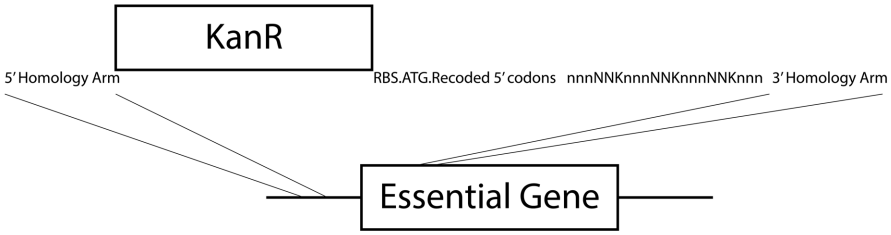
To replace wildtype alleles with corresponding 5' mutant alleles from a mutant library, we used a variation on the methods of Datsenko and Wanner(34). We generated an integration product that consisted (from 5' to 3') of a 5' genomic-targeting homology region, a selectable marker driven by a constitutive promoter, a canonical RBS, recoded DNA of the essential gene (to prevent premature crossover-mediated library excision), the desired degeneracies, and finally a 3' genomic-targeting homology region (see bgg524 for knock-in fragment). This PCR fragment was transformed into freshly prepared electrocompetent MC1061 cells harboring pSIM5. Cells were recovered for one hour in 2YT (as before, the addition of complementing chemical was found to be optional) and then plated on LB agar plates containing the ligand of interest and spectinomycin to select for the marker of the library integration PCR fragment. The resulting library was then subjected to selection and screening.

This method was used to knock-in error prone PCR libraries with the exception that there was no recoding of the ORF and the entirety of the ORF was included on the knock-in fragment. Error prone PCR libraries were generated with the GeneMorph II Random Mutagenesis Kit (Agilent) according to manufacturer's recommendations.

A


5′ Homology Arm
RBS.ATG.
Recoded 5′ codons
nnnNNKnnnNNKnnnNNKnnn
3′ Homology Arm
KanR

B


KanR
5′ Homology Arm
RBS.ATG.Recoded 5′ codons
nnnNNKnnnNNKnnnNNKnnn
3′ Homology Arm
Essential Gene

C

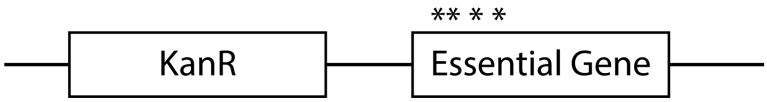

** * *
KanR
Essential Gene

Figure 17. Datsenko-Wanner Style Libraries

A. Any selectable marker is used as template in a PCR amplification step to generate the integration DNA fragment. The 5’ oligo contains homology to the genomic target as well as homology to the antibiotic template for the initial PCR amplification. The 3’ oligo contains (from right to left) 3’ homology to the genomic target, degenerate codons, recoded codons (to prevent premature recombination), and homology to the antibiotic template sequence.

B. The integration fragment (at top) showing the orientation of homology to the genomic target.

C. The resulting genomic locus showing the integrated antibiotic marker and the mutagenized codons (*). Because codons between the mutations and the antibiotic marker were re-coded and share no homology to the wildtype DNA sequence, there is a lower probability of pre-mature recombination leading to the insertion of an antibiotic marker only.

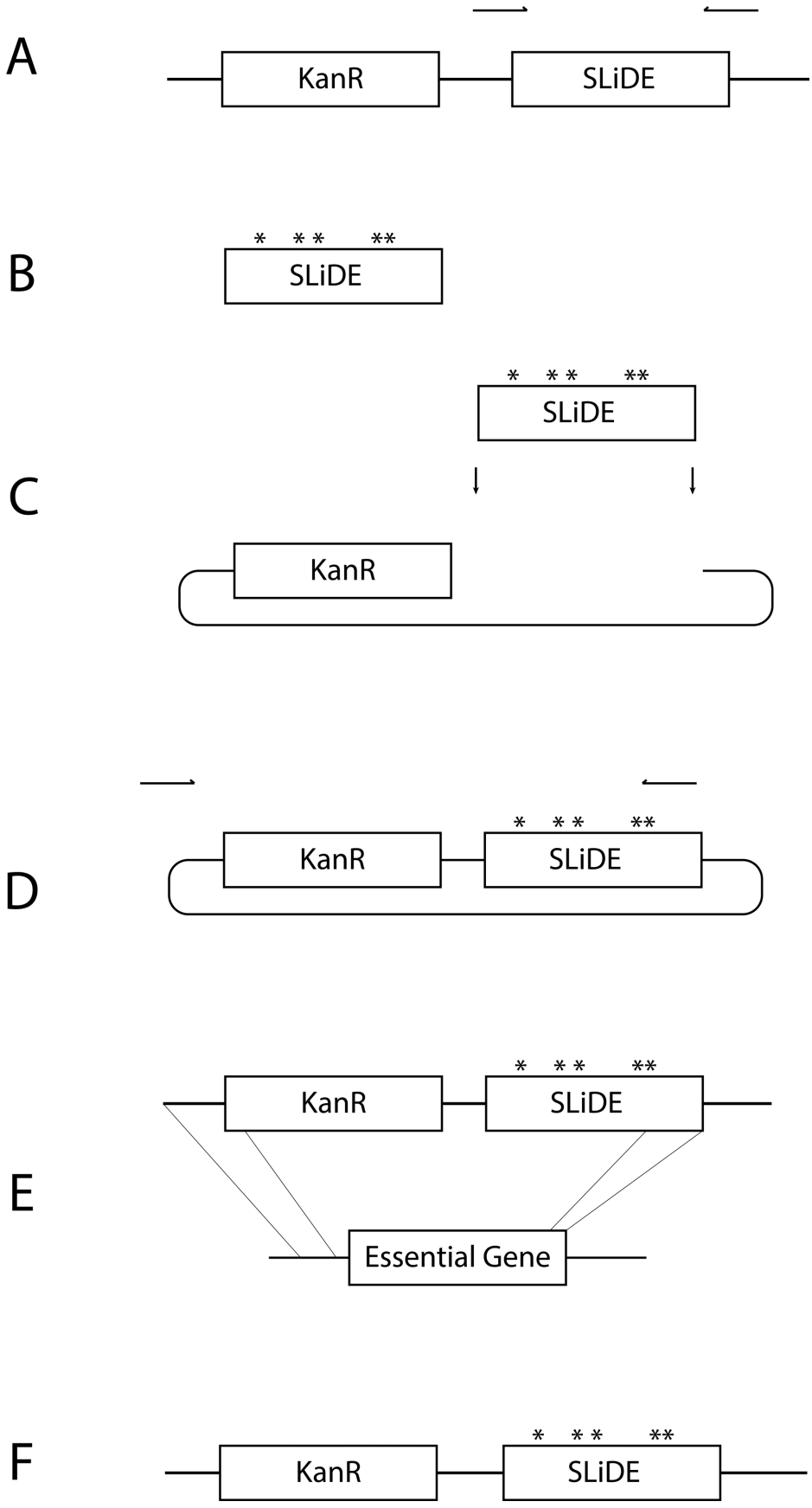

A
KanR
SLiDE
B
SLiDE
C
SLiDE
KanR
D
KanR
SLiDE
E
KanR
SLiDE
Essential Gene
F
KanR
SLiDE

Figure 18. Error Prone PCR Library Knockin Strategy

A. An error prone PCR targets the desired template for amplification.

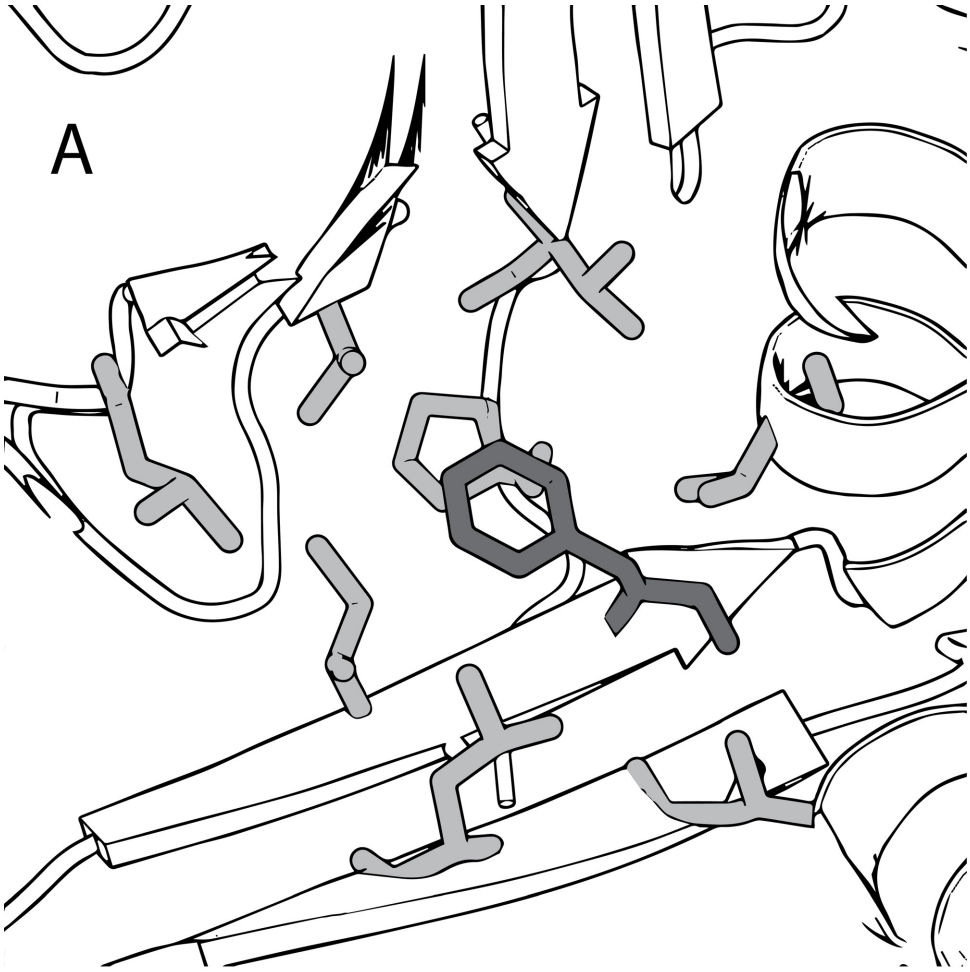
B. The resulting mutagenized DNA fragment is isolated.

C. The mutagenized DNA fragment is sub-cloned.

D. After validating the quality of the library, the plasmid is used as a template to generate an integration fragment by PCR.

E. The resulting integration fragment is knocked into the genome using lambda-red mediated recombination.

F. The resulting SLiDE library can undergo selection and screening.



B

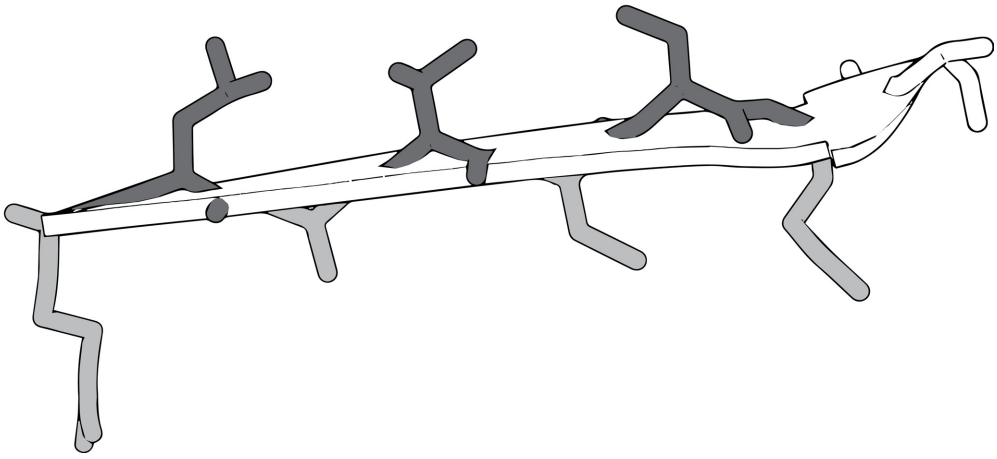


Figure 19. Spherical Library Design Versus Linear Library Approach

A. In the spherical or “dandelion” library approach, a central residue in dark gray would be mutated to glycine, while the surrounding residues in light gray would be randomized.

B. The linear library approach is based on a beta strand in which every other residue points in the same direction. In this case, all of the dark gray residues point up, while the lighter gray residues point down. To the far right, in white, is a proline.

Positive Selection

Positive selections were based simply on viability in the permissive condition. Permissive conditions consisted either of 1M benzothiazole in DMSO used at 1000X or a mixture of small molecules (chemical pool) consisting of 50 mM benzothiazole, 50 mM indole-3-butyric acid, 25 mM indole, and 25 mM 2-aminobenzothiazole dissolved in DMSO as 100X.

Negative Selection

Negative selections were based on Bernard Davis' penicillin technique(21). We found there to be much experimental flexibility(35). Briefly, libraries were re-suspended in media, washed once to remove any complementing chemicals, and re-inoculated into 100 ml of plain 2YT (the restrictive condition) at an OD_{600} of 1-5. Cells were grown at 37 ºC with aeration for 1-2 hours in order to account for phenotypic lag. Penicillin was added at a final concentration of 1mg/ml. The cells were grown for between 5-48 hours. Cells were harvested by centrifugation, washed twice and spread on large LB agar plates containing the desired complementing chemicals for screening.

Phenotypic Screening

A Qpix2 colony picking robot (Genetix) was used for high throughput replica screening. Velveteen was purchased from Stonemountain & Daughter Fabrics in Berkeley, CA. Pintool replicators (384-well and 96-well) were purchased from V&P Scientific. An Epson scanner model V37 was used for scanning replica screening plates. Images were processed using GIMP image processing software.

Screens were variations of replica plating(23). Colonies were either picked by hand or picked by a Qpix2 automated colony picker into 384 well plates. Libraries were replica-gridded by 384-well pintool onto first a restrictive condition screening plate and next onto a permissive condition screening plate. The unmodified replica plating protocol was also used successfully(23).

Replica plate pairs (consisting of mutants gridded on permissive and restrictive plates) were imaged with an Epson V37 office scanner. Using GIMP image editing software, images of scanned plates were overlaid and aligned. The top photo layer was set to 50% opacity. The arithmetic blend mode was set to divide.

Colonies that were viable only in the permissive condition were identified by their colorimetric difference relative to mutants surviving in both conditions.

Preliminary Characterization

Potential SLiDE mutants were re-suspended directly from the screening plate, serially diluted by log10 increments, and replica spotted onto first the restrictive and then the permissive condition. The permissive condition consisted of LB agar plates spiked with individual chemicals. Mutants displaying visible ligand dependencies were sequenced and subjected to higher resolution characterization.

Combining Multiple SLiDE Modules into BL21(DE3)

Previously identified SLiDE strains were used as donor cells to generate P1 phage lysate for transduction. Alternatively, identified mutations were re-integrated into a new host using the Cas9 methodology. BL21(DE3) was first transformed with PKD46-cas9(31) in order to integrate tyrS.GL7 with a single mutagenic oligo as well as an sgRNA targeting WT, non-mutated DNA (bgg539). After clearing the helper plasmid, the resulting strain was P1-transduced with a Δ*tnaA*:FRT-kanR-FRT in order to knockout *tnaA*.

This strain was subsequently transduced with a P1 lysate containing FRT-DHFR-FRT upstream of metG.GL15. The resulting strain was transformed with pFLP2 (in order to remove antibiotic markers associated with metG.GL15 and Δ*tnaA*), outgrown in carbenicillin (100 µg/ml final concentration) for 4 hours, then spread on LB agar plates containing 1 mM benzothiazole and no antibiotic. Viable colonies were screened for antibiotic clearance and sequence confirmed at the relevant loci. This strain was used as the double SLiDE strain.

Subsequently, the double SLiDE strain was P1-transduced with kanR-tagged-pheS.GL2. Resulting colonies were sequenced in order to confirm that they harbored all three SLiDE mutations. These cells were used as triple SLiDE strains.

Characterization of Escape Frequency

Escape frequency was determined by spotting a serial dilution (of log10 increments) of the strain of interest onto LB agar plates of the restrictive condition (consisting of no ligand) and the permissive condition (benzothiazole). Plates

containing DMSO were also spotted to ensure no complementation by the solvent. SLiDE strain escape frequency was continuously monitored during the protein engineering process. The escape frequency of SLiDE strains used in this study (Figure 2D) were assayed on 2-3 separate occasions, using 4 biological replicates derived from re-suspended colonies or liquid cultures. Escape frequency was calculated by dividing the surviving CFU on the restrictive condition by the surviving CFU on 1 mM benzothiazole. The log escape frequencies were used to calculate a mean and standard deviation.

For the double and triple SLiDE strains, an overnight culture was picked into 3 ml 2YT containing 1 mM Benzothiazole and spotted onto 1 mM benzothiazole LB agar plates to validate parental phenotype. The cultures were grown in 24 well blocks at 30ºC with 750 rpm orbital velocity. The next morning, cultures were washed three times with 10% glycerol and re-suspended in 1.5 ml 10% glycerol. From this, 1 ml was spread onto a 220 mm LB agar screening plate containing no complementing ligand and 10 µl of a ten-fold serial dilution was spotted onto a permissive plate, in order to titer CFU plated on the restrictive condition. All plates were grown overnight at 37ºC. The double and triple SLiDE strain escape frequencies were calculated by dividing any escape mutants on the restrictive condition by the total CFU plated as calculated from the serial dilution spotted onto the permissive condition. After the first day, plates were left at room temperature and monitored over the course of 10 days. Each day the total number of colonies was recorded and the escape frequency was calculated for that day by dividing total escape mutants since day one by total CFU plated. The experiment was performed three times for the double mutant and twice for the triple mutant. A total of five biological replicates were used for all experiments from which all mean and standard deviations were calculated. All calculations were performed as described for individual SLiDE strains. In cases where no escape mutants were observed, the maximum limit of detection was calculated as the sum of CFU of all five biological replicates.

Liquid growth assays

A Tecan Sapphire microplate reader was used to measure cell density for growth curves. Growth curves were obtained by re-suspending a fresh colony into 100 µl plain 2YT and depositing 2 µl into a 96-well microtiter plate containing 150 µl 2YT with either benzothiazole (1 mM, 500 µM, 250 µM, 125 µM, 62.5 µM, or 31.25 µM) or without ligand (two replicates). Growth curves were obtained with OD_{600} readings on a Tecan Sapphire every 5 minutes at 37ºC (or at 32ºC for the double and triple mutant) and at an orbital velocity appropriate to the plate. Specific

growth rates (μ) and their corresponding standard deviations were calculated in Excel for 12 biological replicates of the double and triple SLiDE strains at each benzothiazole concentration using the specific growth rate formula(36):

$$\mu = \frac{\ln\left(\frac{x_t}{x_0}\right)}{t}$$

where x_t is OD_{600} at time t_n, x_0 is OD_{600} at time t_{n-1}, t is the time interval $t_n - t_{n-1}$. We used the mean OD of a 6 time point sliding window (30 minutes) in which time points were measured every 5 minutes. The mean and standard deviation were calculated for the maximum growth rates of the 12 biological replicates. This same methodology was used for the SLiDE strain heat map with the following modifications: a 384 well format was used to perform 8 technical replicates at a single ligand concentration of 1000 uM for benzothiazole and indole-3-butyric acid, 500 uM for indole and 2-aminobenzothiazole.

References

1. Schallmey M, Frunzke J, Eggeling L, Marienhagen J (2014) Looking for the pick of the bunch: High-throughput screening of producing microorganisms with biosensors. *Curr Opin Biotechnol* 26:148–154. Available at: http://dx.doi.org/10.1016/j.copbio.2014.01.005.
2. Michener JK, Thodey K, Liang JC, Smolke CD (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab Eng* 14(3):212–222. Available at: http://dx.doi.org/10.1016/j.ymben.2011.09.004.
3. Raman S, Taylor N, Genuth N, Fields S, Church GM (2014) Engineering allostery. *Trends Genet* 30(12):521–528. Available at: http://dx.doi.org/10.1016/j.tig.2014.09.004.
4. Moe-Behrens GHG, Davis R, Haynes K a. (2013) Preparing synthetic biology for the world. *Front Microbiol* 4(5):1–10.
5. Dalby P a. (2011) Strategy and success for the directed evolution of enzymes. *Curr Opin Struct Biol* 21(4):473–480. Available at: http://dx.doi.org/10.1016/j.sbi.2011.05.003.
6. Goldsmith M, Tawfik DS (2012) Directed enzyme evolution: Beyond the low-hanging fruit. *Curr Opin Struct Biol* 22(4):406–412. Available at: http://dx.doi.org/10.1016/j.sbi.2012.03.010.
7. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG (2011) Theoretical and computational protein design. *Annu Rev Phys Chem* 62:129–149.
8. Chica R a., Doucet N, Pelletier JN (2005) Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol* 16(4):378–384.
9. Chen MMY, Snow CD, Vizcarra CL, Mayo SL, Arnold FH (2012) Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng Des Sel* 25(4):171–178.
10. Horowitz NH (1991) Fifty years ago: The neurospora revolution. *Genetics* 127(4):631–635.
11. Beadle GW, Tatum EL (1941) Genetic Control of Biochemical Reactions in Neurospora. *Proc Natl Acad Sci U S A* 27(11):499–506.
12. Woodward R, Stokes L, Larsen, Foster (1943) A Neurospora Assay For Pyridoxine. *J Biol Chem* 150:17–24.
13. Mitchell HK, Houlahan MB (1946) Neurospora. IV. A Temperature-Sensitive Riboflavinless Mutant. *Am J Bot* 33(1):31–35.

14. Gray CH, Tatum EL (1944) X-ray induced Growth factor requirements in Bacteria. *Proc Natl Acad Sci U S A* 30:404–410.

15. Tucker CL, Fields S (2001) A yeast sensor of ligand binding. *Nat Biotechnol* 19(11):1042–1046.

16. Feil R, Metzger D, Chambon P (1997) Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. *Biochem Biophys Res Commun* 757(237):752–757.

17. Tang SY, Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. *Angew Chemie - Int Ed* 50(5):1084–1086.

18. Mandell DJ, et al. (2015) Biocontainment of genetically modified organisms by synthetic protein design. *Nature* 518(7537):55–60. Available at: http://dx.doi.org/10.1038/nature14121.

19. Rovner AJ, et al. (2015) Recoded organisms engineered to depend on synthetic amino acids. *Nature* 518(7537):89–93. Available at: http://dx.doi.org/10.1038/nature14095.

20. Baba T, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2.

21. Davis BD (1949) Isolation of biochemically deficient mutants of bacteria by penicillin. *J Am Chem Soc* 35(1):1–10.

22. Marisch K, Bayer K, Cserjan-Puschmann M, Luchner M, Striedner G (2013) Evaluation of three industrial Escherichia coli strains in fed-batch cultivations during high-level SOD protein production. *Microb Cell Fact* 12(58):1–11. Available at: Microbial Cell Factories.

23. Lederberg J, Lederberg EM (1952) Replica plating and indirect selection of bacterial mutants. *J Bacteriol* 63(3):399–406.

24. Keseler IM, et al. (2013) EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Res* 41(Database issue):D605–D612.

25. Deckert K, Budiardjo SJ, Brunner LC, Lovell S, Karanicolas J (2012) Designing allosteric control into enzymes by chemical rescue of structure. *J Am Chem Soc* 134(24):10055–10060.

26. Levay-young B, et al. (2013) Library Format for Bioengineering Maximizing Screening Efficiency through Good Design. *Genet Eng Biotechnol News* 33(8):1–3.

27. Wright O, Delmans M, Stan G-B, Ellis T (2014) GeneGuard: a Modular Plasmid System Designed for Biosafety. *ACS Synth Biol* (iii):307–316. Available at: http://pubs.acs.org/doi/abs/10.1021/sb500234s\nhttp://www.ncbi.nlm.nih.gov/pubmed/24847673.

28. Li G, Young KD (2013) Indole production by the tryptophanase TnaA in escherichia coli is determined by the amount of exogenous tryptophan. *Microbiol (United Kingdom)* 159(2):402–410.

29. Douthwaite B, Keatinge JDH, Park JR (2001) Why promising technologies fail: The neglected role of user innovation during adoption. *Res Policy* 30(5):819–836.

30. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* 98(12):6742–6746.

31. Peng Z, Richardson S, Robinson D, Deutsch S, Cheng J (2014) *Genome Editing in Escherichia coli with Cas9 and synthetic CRISPRs*.

32. Datta S, Costantino N, Court DL (2006) A set of recombineering plasmids for gram-negative bacteria. *Gene* 379:109–115.

33. Esvelt KM, et al. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 10(11):1116–21. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3844869&tool=pmcentrez&rendertype=abstract.

34. Datsenko K a, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 97(12):6640–6645.

35. Rossi JJ, Berg CM (1971) Differential recovery of auxotrophs after penicillin enrichment in Escherichia coli. *J Bacteriol* 106(2):297–300.

36. Stanbury PF, Whitaker A, Hall SJ (2013) *Principles of Fermentation Technology* (Elsevier Science) Available at: https://books.google.com/books?id=CW8vBQAAQBAJ.