Title
Design of Biological Systems for Tools and Applications

Permalink
https://escholarship.org/uc/item/2t54j0fz

Author
Kittleson, Joshua

Publication Date
2012

Peer reviewed|Thesis/dissertation
Design of Biological Systems for Tools and Applications

By

Joshua Tyler Kittleson

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 Adam Arkin
Professor Daniel Portnoy
Professor Francis Szoka

Fall 2012
Abstract

Design of Biological Systems for Tools and Applications

by

Joshua Tyler Kittleson

Joint Doctor of Philosophy in Bioengineering with UCSF

University of California, Berkeley

Professor J. Christopher Anderson, Chair

Early efforts to genetically engineer biological systems were restricted to relatively crude cut and paste operations, with the sequence of even model organisms’ genomes inaccessible for detailed study. Exponential improvement in our ability to both read and write DNA sequences has dramatically altered the landscape of genetic engineering, rendering de novo synthesis of entire genomes possible [1]. However, our ability to design functional systems has not kept pace with our ability to fabricate them. The field of synthetic biology seeks to address this shortcoming and provide a theoretical framework for rationally manipulating biological systems. Of particular interest is exploration of how core engineering principles such as modularity, abstraction, and standardization can be applied to the development of genetic systems. Despite considerable recent progress toward answering that question, imperfect and incomplete knowledge about biological systems necessitates exploration of numerous variants to craft even relatively simple systems. In this work, we first explore development of biological systems that make engineering biology easier, then pursue development of a complex system for a demanding application to probe the limits of our design capability, and conclude with a discussion of key challenges in designing biological systems and how to address them.

Motivated by the observation that tuning the expression of system components is often essential to achieving desired functions, we begin by describing development of a system for rapidly prototyping genetic constructs to determine an optimal expression level. Although several methods exist for manipulating the transcription or translation of a desired gene, they require fabrication of numerous distinct variants. To work around this limitation, we instead explore manipulation of plasmid copy number and develop a set of strains that are capable of maintaining the same plasmid at a desired level in the range of 1-250 copies per cell. We demonstrate that this system can be applied to rapidly find the optimal expression level for a model biosynthetic pathway, regardless of the transcriptional activity of the pathway. Recognizing that fabricating and assaying distinct variants is in some cases essential, we then examined the problem of enabling high-throughput manipulation of DNA. Existing methods are limited by difficult to automate operations, such as centrifugation, application of a vacuum, or gel electrophoresis. Development of alternative methods that only require liquid handling operations would enable utilization of very high throughput automation platforms. We therefore engineer a P1 phage-based system for transferring plasmids between cells using only liquid
handling operations. After establishing exogenous control of phage lysis through addition of a small molecule inducer, we incorporate phage cis elements that enable desired DNA to be transferred 1600-fold more efficiently than other cellular DNA. The system is capable of transferring large libraries of DNA up to 25 kilobases in length at small volumes. This provides an important first step toward development of a highly automatable suite of tools for biological engineering.

Biological engineers ultimately want to manipulate biological systems to solve human specified problems. Although a number of potential applications of synthetic biology have been described [2–4], most projects rely on relatively simple designs to achieve a desired outcome. To identify bottlenecks in the development of more complex systems, we investigate engineering a laboratory strain of *E. coli* to localize to, invade, and kill cancer cells. After attempting to transfer capsular polysaccharide synthesis clusters to a laboratory strain to improve bacterial survival in an animal model, we generate a strain capable of delivering a cytotoxic ribonuclease to and subsequently killing cultured cancer cells. Although the strain did not inhibit tumor growth in an animal model, the lessons learned during the execution of this and other projects guided our thinking about current bottlenecks and the path forward. We conclude with a discussion of the potential for modular design to enable routine development of complex biological systems, and identify six failure modes that hinder current efforts. By architecting software to codify existing biological knowledge and investigating the basis of important phenomena such as load and stress, we can harness the power of improved DNA sequencing and synthesis capabilities to build novel, useful biological systems.
# Table of Contents

Table of Contents................................................................. i
Table of Figures........................................................................ iii
Table of Tables.......................................................................... iv
Acknowledgements.................................................................. v

## Chapter 1 – Introduction
1.1 – Synthetic biology and biological engineering.................... 1
1.2 – Expression level optimization........................................... 1
1.3 – High throughput screening to overcome incomplete information........ 2
1.4 – Application development............................................... 2
1.5 – Remaining challenges.................................................... 3

## Chapter 2 – Rapid Optimization of Gene Dosage in *E. coli* using DIAL Strains
2.1 – Abstract........................................................................... 4
2.2 – Background....................................................................... 4
2.3 – Results and Discussion.................................................... 5
  2.3.1 – ColE2 as a Trans-Activated Origin.............................. 5
  2.3.2 – Orthogonality of R6K and ColE2................................. 6
  2.3.3 – Construction of DIAL Strains...................................... 6
  2.3.4 – Characterization of DIAL Strains: Copy Number, Cell-to-Cell Variability, and Stability........................................ 7
  2.3.5 – Optimization of Violacein Expression.......................... 9
2.4 – Conclusions....................................................................... 10
2.5 – Methods............................................................................ 11
  2.5.1 – Media.......................................................................... 11
  2.5.2 – Plasmids...................................................................... 12
  2.5.3 – Genomic RBS Library Construction................................ 12
  2.5.4 – Plasmid Copy Number Estimation by qPCR...................... 13
  2.5.5 – Flow Cytometry........................................................... 13
  2.5.6 – Stability Analysis......................................................... 13
2.6 – Acknowledgements.......................................................... 13

## Chapter 3 – Scalable Plasmid Transfer using Engineered P1-based Phagemids
3.1 – Abstract........................................................................... 14
3.2 – Introduction...................................................................... 14
3.3 – Results and Discussion.................................................... 16
  3.3.1 – Transcription of *coi* Triggers the P1 Lytic Cycle........... 16
  3.3.2 – *cin* is a cis Element that Improves Phagemid Transduction..................................................... 17
  3.3.3 – Phagemids Exhibit Robust and Faithful Transduction........ 18
3.4 – Summary and Conclusions................................................. 21
3.5 – Methods............................................................................ 22
  3.5.1 – Plasmids, Strains, Phage, and Growth Media.................. 22
  3.5.2 – Generation of Phage Lysates........................................ 23
  3.5.3 – Transduction of Phage Lysates..................................... 23
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1.</td>
<td>qPCR estimation of plasmid copy number</td>
<td>6</td>
</tr>
<tr>
<td>Figure 2.2.</td>
<td>Flow cytometric analysis of cell-to-cell variation</td>
<td>8</td>
</tr>
<tr>
<td>Figure 2.3.</td>
<td>Stability of plasmids in DIAL strains</td>
<td>9</td>
</tr>
<tr>
<td>Figure 2.4.</td>
<td>Optimization of violacein production</td>
<td>10</td>
</tr>
<tr>
<td>Figure 2.5.</td>
<td>Schematic of templates constructed for genomic integration</td>
<td>12</td>
</tr>
<tr>
<td>Figure 3.1.</td>
<td>Genetic interaction of phage lysogen and phagemid</td>
<td>16</td>
</tr>
<tr>
<td>Figure 3.2.</td>
<td>Expression of <em>coi</em> induces lysis</td>
<td>17</td>
</tr>
<tr>
<td>Figure 3.3.</td>
<td>Impact of <em>cis</em> elements on packaging</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3.4.</td>
<td>Impact of volume on phagemid efficiency</td>
<td>19</td>
</tr>
<tr>
<td>Figure 3.5.</td>
<td>Impact of phagemid size on efficiency</td>
<td>20</td>
</tr>
<tr>
<td>Figure 3.6.</td>
<td>Library transfer using the phagemid</td>
<td>21</td>
</tr>
<tr>
<td>Figure 4.1.</td>
<td>Strategies for manipulation of capsular polysaccharides</td>
<td>28</td>
</tr>
<tr>
<td>Figure 4.2.</td>
<td>Survival of encapsulated bacteria in a mouse</td>
<td>29</td>
</tr>
<tr>
<td>Figure 4.3.</td>
<td><em>In vitro</em> killing of tumor cells</td>
<td>32</td>
</tr>
<tr>
<td>Figure 4.4.</td>
<td>Bacterial treatment of tumors in mice</td>
<td>33</td>
</tr>
<tr>
<td>Figure 5.1.</td>
<td>Failure modes of genetically engineered systems</td>
<td>39</td>
</tr>
</tbody>
</table>
Table of Tables

Table 2.1. RBSs of repA and pir expression variants........................................................... 7
Table 2.2. Oligos used for library construction........................................................................ 12
Table 4.1. In vitro assays of capsular polysaccharides.......................................................... 30
Table 4.2. Bacterial Tumor Colonization and Plasmid Maintenance...................................... 31
Table 6.1. Basic Parts............................................................................................................. 45
Table 6.2. Composite Parts................................................................................................... 46
Table 6.3. Plasmid Backbones.............................................................................................. 48
Table 6.4. Strains.................................................................................................................. 48
Acknowledgements

I am deeply grateful to the dedicated individuals who have either directly contributed to my development as a scientist and the work presented here or supported me as I navigated through that process. First and foremost, I want to thank my advisor, Chris Anderson, for his thoughtful guidance and tutelage over the past six and half years. From teaching me how to perform the technical operations required of a molecular biologist to providing insight into developing informed plans for tackling complex problems and broadening my perspective on what it means to be a biological engineer, his constant and patient direction has been an invaluable and formative part of my experience. I have also been fortunate to have had significant interactions with both John Dueber and Adam Arkin, whose insight and ideas have energized my thinking about synthetic biology. The other members of the Anderson lab have been a pleasure to work with and be around, with Tim Hsiau and Jin Huh in particular facilitating this work. The ecosystem of other synthetic biology labs, including the Arkin and Dueber labs, has provided an excellent academic environment, such as numerous thoughtful conversations with Weston Whitaker. Sherine Cheung effectively implemented several of the experiments reported in Chapter 2, while Will DeLoache was instrumental in kicking off the phagemid research reported in Chapter 3 and Lindsey Saldin put a lot of time and effort into the capsular polysaccharide material presented in Chapter 4. Gabriel Wu both focused the discussion presented in Chapter 5 and helped provide some balance to my experience as a graduate student. I would also like to thank the Joint Graduate Group in Bioengineering, SynBERC, the National Science Foundation, and the Siebel Scholars for funding this work.

Finally, I want to thank my family for their support these past several years. Though somewhat convinced that my work might cause a zombie apocalypse, my parents and siblings have helped keep me focused and productive. My wife, Jennifer Hogan, has persevered through a long-distance relationship, an erratic schedule, and more talk about moving clear liquids around than any sane person should be subjected to, and has been nothing but supportive and patient throughout. I’ve been very fortunate to have been surrounded by understanding and encouraging people.
Chapter 1 – Introduction

1.1 – Synthetic Biology and Biological Engineering

Biological systems are highly complex, with interesting and relevant phenomena occurring across several time and space scales. Further, such systems can behave stochastically, are frequently out of equilibrium, and are imperfectly self-replicating [5]. Biologists have been trying to observe, describe, and understand these complex natural biological systems for decades, despite only recently having acquired the technical capability to observe specific sub-cellular events in real-time in a living system [6]. Out of this ongoing work to develop a molecular level understanding of how cells and organisms function, some investigators have turned toward altering biological systems to achieve a human specified purpose. Fortunately for such engineers, the intricate molecular dance of living cells is largely orchestrated by DNA, a relatively easy to observe central information repository. Early manipulations of DNA were limited to simple cut and paste operations [7], or mutation of a specific DNA sequence [8]. Even this limited capacity for manipulation, however, was sufficient to enable development of the biotechnology industry [9]. Since the inception of recombinant DNA, our ability to both read and write the genetic code has expanded exponentially [10]. The field of synthetic biology has emerged in recognition of the fact that we lack a formal theoretical framework for selecting what DNA sequence to write to achieve a specified goal [11]. In particular, we have not yet established how key engineering concepts such as abstraction, standardization, and modularity can be applied to the design of complex biological systems. Early work in the field focused on demonstrating our capability to build interesting genetic systems, showing that we have sufficient control over basic cellular processes to rationally generate a DNA sequence that elicits a desired set of non-native behaviors [2]. Subsequent work has turned toward finding applications for these new kinds of genetic circuits [2-4]. Importantly, it remains clear that in the absence of a complete understanding of biological behaviors, examination of numerous variants is required to achieve even modest engineering goals.

In this work, we focus on applying important ideas from synthetic biology to the development of tools that make engineering biology easier, pushing on the limits of our design capability, and reflecting on remaining challenges in engineering biological systems. First, we develop a system for rapidly prototyping genetic constructs to determine an optimal expression level. By enabling different cell lines to maintain the same plasmid at different copy numbers, we reduce the necessary operations down to a single simple transformation. Second, we engineer a phage-based system for transferring plasmids between cells. Because the method requires only liquid handling operations, it enables automation using standard, extremely high throughput robotics platforms. After developing these tools, we try to engineer a laboratory strain of bacteria to localize to, invade, and destroy tumor cells. The complexity of the pathways and interactions required challenges our ability as biological engineers. Finally, we draw from these and other experiences to outline limitations in our ability to engineer biological systems. By highlighting specific failure modes and suggesting potential solutions, we offer a path forward.

1.2 – Expression Level Optimization
Biological engineers frequently need to optimize the expression level of one or more genes, whether for improving the flux through a metabolic pathway [12], or for tuning the behavior of a genetic circuit [13]. Typically, this has been approached by cloning a number of distinct plasmid variants. However, development of a more efficient strategy would accelerate the engineering process. Chapter 2 describes the development of a novel biological tool for rapidly scanning through almost two orders of magnitude of expression levels using only a single plasmid and a simple protocol. A crucial part of engineering the system was replacing the native regulation of two proteins responsible for controlling plasmid replication with synthetic constitutive promoters. Because biological engineers largely draw from existing biological functions and native systems usually have transcriptional and/or translational regulation, the need to re-engineer regulatory signals is a common practice in synthetic biology. This process, referred to as refactoring, is also employed in chapter 3 to link a biochemical event to a human specified signal, and to a lesser extent in chapter 4 to control an important cellular phenotype. Chapter 5 expands on the idea to discuss how understanding the context and regulation of a genetic system are critical to system function.

1.3 – High throughput screening to overcome incomplete information

Because biological engineers typically have insufficient information to specify the precise components, configuration, and regulation necessary to implement a desired system as basepair level precision, it is usually necessary to construct and test numerous variants. In chapter 2, for example, hundreds of ribosome binding site variants are tested to find the desired behavior, and in chapter 4, several proteins are tested for their ability to cause cytotoxicity in cancer cells. The field would therefore benefit from improvements to the throughput of the fundamental operations required to construct and test new designs. Chapter 3 argues that conversion of routine procedures to liquid handling operations is essential to achieving a transformative increase in the throughput of construct fabrication and analysis. It proceeds to describe the development of a phage-based system for transferring plasmids between cells using only liquid handling operations. In order to improve the efficiency of the system, millions of variants are competed against each other, utilizing the fact that more efficient systems amplify faster than less efficient ones. For other design efforts, where it may be more difficult to link a growth advantage to a desired behavior, development of these kinds of tools is essential. This limitation is further explored in chapter 5, which notes that until we can develop a more comprehensive way of predicting cellular behaviors, we have to resort to this type of trial and error.

1.4 – Application development

Ultimately, development of biological tools is intended to facilitate engineering of biological systems to address a problem. Chapter 4 presents our efforts to engineer tumor killing bacteria, a problem selected to explore the limitations of our current engineering capabilities. We first manipulate large biosynthetic operons to confer a laboratory strain of E. coli with a controllable way of expressing protective polysaccharide capsules. We then engineer strains to deliver cytotoxic proteins to cancer cells, and demonstrate that they are capable of killing cultivated tumor cells. Challenges encountered during the engineering process prompted
discussion of several failure modes in chapter 5, including syntactic error, semantic errors, context dependence, and concerns about load and stress.

1.5 – Remaining Challenges

In the course of implementing the projects described in chapters 2-4, we observed that failures during development of a biological system generally fall into one of a handful of categories, and that we could take steps to help minimize those failures. Chapter 5 describes the potential of applying a modular engineering strategy, and then considers six general types of system failure. The failure modes encompass a broad spectrum of problems, ranging from simple human design error to population heterogeneity. We then describe how computer aided design can help mitigate some errors, while others can only currently be addressed by trial and error. Finally, chapter 5 argues that continued development of DNA synthesis and manipulation techniques, coupled to further research into the causes of cellular load and stress, will enable better a priori design.
Chapter 2 – Rapid Optimization of Gene Dosage in *E. coli* using DIAL Strains

Authors:

Joshua T Kittleson  
Department of Bioengineering, University of California, Berkeley CA, United States

Sherine Cheung  
Department of Bioengineering, University of California, San Diego, USA

J. Christopher Anderson (corresponding author: jcanderson@berkeley.edu)  
Department of Bioengineering, University of California, Berkeley CA, United States  
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley CA, United States

Reprinted from *Journal of Biological Engineering*, JT Kittleson, SCheung, JC Anderson, “Rapid optimization of gene dosage in *E. coli* using DIAL strains”, Vol. 5:10, (2011) under the terms of the Creative Commons Attribution License

2.1 – Abstract

Engineers frequently vary design parameters to optimize the behaviour of a system. However, synthetic biologists lack the tools to rapidly explore a critical design parameter, gene expression level, and have no means of systematically varying the dosage of an entire genetic circuit. As a step toward overcoming this shortfall, we have developed a technology that enables the same plasmid to be maintained at different copy numbers in a set of closely related cells. This provides a rapid method for exploring gene or cassette dosage effects. We engineered two sets of strains to constitutively provide a *trans*-acting replication factor, either *Pl* of the R6K plasmid or RepA of the CoIE2 plasmid, at different doses. Each DIAL (different allele) strain supports the replication of a corresponding plasmid at a constant level between 1 and 250 copies per cell. The plasmids exhibit cell-to-cell variability comparable to other popular replicons, but with improved stability. Since the origins are orthogonal, both replication factors can be incorporated into the same cell. We demonstrate the utility of these strains by rapidly assessing the optimal expression level of a model biosynthetic pathway for violacein. The DIAL strains can rapidly optimize single gene expression levels, help balance expression of functionally coupled genetic elements, improve investigation of gene and circuit dosage effects, and enable faster development of metabolic pathways.

2.2 – Background

Optimizing desired outcomes by varying a design parameter is a staple of almost every engineering field, from mechanical engineers tweaking blade angles on a wind turbine to civil engineers altering the timing of traffic lights. Similarly, genetic engineers alter gene expression levels to optimize some desirable phenotype. Strong overproduction of single proteins can
impose a metabolic burden on E. coli, and often a lower expression level leads to improved phenotype [14]. In multi-subunit proteins and genetic circuits, expression of particular proteins often needs to be balanced for proper function (e.g. [15],[16], and [17]). Extensive work has established methods for achieving expression of a gene or operon at a particular level, including control of transcription using standard promoter sets [18], modulation of RNA processing [19], and control of translation through ribosome binding site (RBS) manipulation [20]. However, using these tools to investigate the desired expression level of a single gene or operon requires cloning for each level to be tested. Using inducible promoter systems to probe multiple expression levels can rapidly determine an approximate desired expression level, but does not provide a genetically encoded solution, which can be useful for downstream applications. For optimizing multi-operon constructs, fewer tools exist. Generating large numbers of repeats in the genome is labor intensive[21], while strategies that increase plasmid copy number upon induction provide a narrow range of copy numbers [22], cause runaway replication [23, 24], or are incompatible with constructs using the common P_{BAD} promoter [25]. To address these shortcomings and allow researchers to explore the effect of copy number on genetic devices, we have exploited an underutilized control point: plasmid copy number.

Genetic engineers working in E. coli are blessed with a wide range of plasmid systems and plasmid copy numbers to choose from, ranging from single copy BACs to ~500 copy pUC plasmids[26]. However, to take advantage of copy number differences, each gene or device of interest has to be cloned into different plasmids, necessarily changing the local genetic context along with the copy number. A notable exception to this rule is the gamma origin of the R6K plasmid, which requires the trans-acting Pi protein to initiate replication [27]. Different alleles of pir integrated into the E. coli genome are known to support R6K plasmids at different copy numbers (e.g. pir+ and pir116 [28]), meaning that the genetic context on the plasmid itself remains unaltered. Similarly, the orthogonal replicon of ColE2-P9 also uses a trans-acting factor, RepA, to support the origin of replication [29].

In this work, we generate two sets of strains bearing different alleles (DIAL) of pir or repA to support the same plasmid at a wide range of copy numbers. We then characterize the copy number, cell-to-cell variability, and stability of plasmids in both sets of DIAL strains. We illustrate their utility on a model system by examining expression of the violacein biosynthesis pathway [30] at different copy numbers. The results demonstrate that artificially re-regulating replication factor expression from the genome can produce stable plasmid copy numbers, that phenotype varies with copy number, and that DIAL strains can accelerate development of genetic devices.

2.3 – Results and Discussion

2.3.1 – ColE2 as a Trans-Activated Origin

A previous report[31] identified a minimal 32 bp region of ColE2 sufficient to support replication of a plasmid when repA is provided in trans. We first recapitulated this behaviour by transforming a plasmid bearing the ColE2 minimal origin (BBa_J72203-BBa_J72158) into cells constitutively expressing the trans-acting repA gene from a plasmid (MC1061 + BBa_J72109-BBa_J72182). Although we observed transformants, they exhibited small colony morphologies. Presuming this observation to reflect plasmid instability (as suggested by data in [32]), we repeated the experiment using a larger 470 bp fragment of the ColE2 plasmid as the origin
(BBa_J72203-BBa_J72161) in the hope that any context dependent influence on the minimal origin would be eliminated, or that non-essential factors contributing to robustness would be captured. This yielded colonies with morphologies indistinguishable from untransformed cells (data not shown).

2.3.2 – Orthogonality of R6K and ColE2

We next examined if ColE2 and R6K were orthogonal origins of replication. Plasmids bearing either an R6K or ColE2 origin of replication were transformed or cotransformed into cells expressing pir, repA, or both. We observed colonies only when a plasmid was transformed into cells possessing the cognate replication factor, as expected.

![Figure 2.1 – qPCR estimation of plasmid copy number.](image)

**Figure 2.1 – qPCR estimation of plasmid copy number.** JTK160 (repA) and JTK164 (pir) variants were transformed with BBa_J72205-BBa_J72048 and BBa_J72206-BBa_J72048, respectively. JTK160J and JTK164J were also transformed with BBa_J72202-BBa_J72048 (p15a origin) and BBa_J72207-BBa_J72048 (pUC origin). Two biological replicates of each sample were prepared and analyzed using qPCR. Error bars represent standard deviation.

2.3.3 – Construction of DIAL Strains

To generate cells expressing diverse levels of the trans-acting factors Pir and RepA, we integrated expression cassettes with randomized ribosome binding sites (RBSs) into the genome, thereby creating strain sets JTK160 and JTK164 for pir and repA, respectively. We subsequently visualized copy number variation by transforming the libraries with reporter plasmids constitutively expressing sfGFP [33] (BBa_J72111-BBa_J72183 or BBa_J72112-BBa_J72183).
After a preliminary fluorescence analysis of 380 clones of each type (data not shown), 24 were selected for further investigation. After a second round of fluorescence measurement, 10 variants of each type that spanned the range of observed fluorescence levels were chosen for full characterization. The sequences of the selected RBSs are reported in Table 2.1.

### 2.3.4 – Characterization of DIAL Strains: Copy Number, Cell-to-Cell Variability, and Stability

We characterized three important properties of plasmids in the DIAL strains: copy number, cell-to-cell variation, and stability. To estimate the copy number supported by each strain, we employed qPCR to examine plasmid content both at mid log and at stationary phase (Figure 2.1). Based on this analysis, a ColE2 plasmid in the DIAL strains spans the range of ~1-60 copies per genome equivalent, while an R6K plasmid in the DIAL strains spans the range of ~5-250 copies per genome equivalent. This covers nearly the entire range of reported plasmid copy numbers, from single copy to nearly pUC levels. We observed that the pUC plasmid exhibited 4-5 fold increased copy numbers at stationary phase, while the p15a, R6K, and ColE2 plasmids showed ~2-fold or lower changes.

#### Table 2.1 – RBSs of repA and pir expression variants

<table>
<thead>
<tr>
<th>Strain</th>
<th>Variant</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>JTK160 (repA)</td>
<td>A</td>
<td>TCTAGAATACCTGATG</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>GTGAGAAGTGAACGTG</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>CCGAGAACGTAGGATG</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>GGCAGAAGTGAATG</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>GTGAGAAAGCTCTGTG</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>CTCAGAACCAGAATG</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>GTGAGAATGCTTTATG</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>TTGAGAAACAGATG</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>GGGAGAAAGATGGGTG</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>GGGAGAAACAAAAATG</td>
</tr>
<tr>
<td>JTK164 (pir)</td>
<td>A</td>
<td>GCTGGAAACAGGTTGTG</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>TAAGGAATAGGTGTG</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>GGGGGAAGGGCATGTG</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>TTGGGAACATTCATG</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>TCCGGAAGACTAGGTG</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>GTCGGAAAGGCTGTG</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>GTGGGAATAGAATATG</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>ACGGGAATGTAACGTG</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>ACTGGAAACATAGGTG</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>ATCGGAAACATAGGTG</td>
</tr>
</tbody>
</table>

Start codons are underlined

We next examined sfGFP expression in samples of each of the cells by flow cytometry (Figure 2.2) to determine cell-to-cell variability. The ColE2 and R6K plasmids generally exhibit similar distributions to p15a and pUC origins. Strains JTK160I and JTK164E, which have mean
expression levels within 25% of p15a levels, have coefficients of variance that fall within 25% of p15a levels. Similarly, JTK160J and JTK164I, which have mean expression levels within 25% of pUC levels, have coefficients of variance that fall within 25% of pUC levels. It is unclear if the relatively high coefficients of variance for JTK160A and JTK164A are a result of noise at low fluorescence (as evidenced by the very high MC1061 coefficients of variance) or due to true variance in plasmid copy number or GFP expression.

Figure 2.2 – Flow cytometric analysis of cell-to-cell variation. JTK160 strains expressing repA (A) or JTK164 cells expressing pir (B) bearing GFP expressing plasmids (BBa_J72111-BBa_J72185 or BBa_J72112-BBa_J72185, respectively) were analyzed by flow cytometry. MC1061 alone or bearing p15a (BBa_J72202-BBa_J72185) or pUC (BBa_J72109-BBa_J72185) plasmids expressing GFP were also analyzed. Data is presented as forward scatter versus GFP fluorescence, with the mean and coefficient of variance listed below.

Finally, we monitored the stability of plasmids in mid-level expression variants of both pir and repA after 100 generations without selection (Figure 2.3). Both pir and repA variants
exhibited high stability, losing the plasmid in only 5.2% or 0.5% of cells, respectively. These numbers fall between the stability of the control p15a and pUC plasmids, which lost the plasmid in 23.5% or .25% of cells, respectively.

2.3.5 – Optimization of Violacein Expression

As a simple demonstration of the utility of the DIAL strains, we optimized the expression level of the violacein biosynthesis operon, VioABCDE. While moderate production levels of the deeply purple metabolite are tolerated in *E. coli*, high levels can be toxic or cause instability [30]. We cloned the operon behind both a weak and a strong constitutive promoter to illustrate the flexibility afforded by controlling copy number. Figure 2.4 shows the colonies that result from transforming both weakly and strongly expressed operons (BBa_J72188 and BBa_J72189, respectively) into all of the DIAL strains. Production of violacein, as indicated by purple coloration, clearly increases as copy number increases, up to some threshold level. Beyond that threshold, colonies begin to grow at reduced rates or not at all. The large colonies in the high copy strains (JTK160G, JTK160I, and JTK164H) with a strong promoter were sequenced and confirmed to be escape mutants in which a fragment of the strong constitutive promoter has recombined out, demonstrating the risk of overstressing the cells.

Figure 2.3 – Stability of plasmids in DIAL strains. JTK165E1 (*pir repA*) bearing BBa_J72202-BBa_J72183 (p15a origin), BBa_J72207-BBa_J72183 (pUC origin), BBa_J72111-BBa_J72183 (ColE2 origin), or BBa_J72112-BBa_J72183 (R6K origin) was serially propagated for 100 generations and then analysed for plasmid loss. Four biological replicates of each plasmid were analysed, and error bars represent standard deviation.
2.4 – Conclusions

We have developed and characterized two sets of strains that support the R6K and ColE2 origin of replication at a wide range of different copy numbers enabling rapid exploration of gene and circuit dosage. To accomplish this, we placed the trans-acting replication factors from each replicon under artificial transcriptional regulation in the genome, leaving only the origins of replication on the plasmids themselves. Although negative feedback relying on elements 5’ of the trans-acting factor open reading frame has been implicated as one factor helping to maintain stable copy numbers in both ColE2 [34] and R6K [27], we found that engineered cells stably maintained plasmid copy numbers despite removal of all 5’ regulatory elements. This is consistent with the existence of additional feedback mechanisms, as has been suggested for both R6K [35] and ColE2 [36].

To generate copy number diversity in the DIAL strains, we created a library of RBS variants of the trans-acting replication factor in the genome. Although other strategies, such as the use of an inducible promoter or a library of promoters, could also have achieved diverse levels of trans-acting factor expression, varying the RBS enables compatibility with genetic circuits employing any promoter and maintains a consistent noise profile across strains due to stochastic transcription effects [37]. We employed a novel mechanism of generating RBS libraries in the genome: lambda red based integration of Splicing by Overlap Extension (SOEing) [38] PCR
products. Multiplex automated genome engineering [39] has also been employed for creating libraries of modified genomic RBSs. While that process is a powerful method for modifying genes already in the genome, this work required simultaneous modification and introduction into the genome of an exogenous gene. In such cases, PCR based integration is an excellent option for library construction, particularly where a relatively small number of variants (<10,000) is sufficient to isolate the desired functionality.

The DIAL strains are the first tool capable of systematically varying genetic circuit dosage without altering the local genetic context. Previous studies examining the impact of circuit dosage in prokaryotes have been largely theoretical (e.g. [40, 41]), and in eukaryotes focus only on low (~1-2) copy numbers (e.g. [42]). Because the theoretical predictions suggest that circuit dosage has a significant impact on the function of some genetic circuits, it is important to empirically verify the robustness or fragility of different circuit architectures. Using the DIAL strains, network behaviour and expression noise can be rapidly assessed at a wide variety of different circuit dosages.

Of great practical use, the DIAL strains offer a rapid, facile mechanism for determining desired expression levels, making it a tool with broad applicability in genetic engineering. The trivial operation of transforming the same plasmid into different strains is sufficient to provide information on the maximum tolerated expression level for a given protein, pathway, or circuit, and screening of viable colonies reveals the optimal expression level for a desired phenotype. We demonstrated this simple capability by optimizing expression of the violacein biosynthesis pathway, which in excess produces moderate toxicity in *E. coli*. Regardless of whether that starting point was a weakly or a strongly expressed operon, deeply purple yet healthy cells were isolated when matched with the appropriate strain. Knowing the optimal gene dosage can be leveraged to change the context of a gene or operon without altering the phenotype. Since the copy number is known, any change in protein dosage resulting from changing the context of the system (such as by integration into the genome) can be compensated for by using existing tools such as the RBS calculator [20] or a set of standard promoters [18].

Importantly, the ColE2 and R6K origins are orthogonal and can co-exist in the same cell, and the two sets of DIAL strains were designed to enable ready combination of both trans factors into a single strain by P1 transduction. Having a single set of cells with both orthogonal origins allows both the relative and absolute levels of two genes or sets of genes to be optimized by, for example, co-transformation into a pool of competent cells. Although R6K has already seen widespread use because of its ability to split into trans and cis elements, having a variety of copy number variants available for both R6K and ColE2 provides an even more powerful toolset for expression level optimization and balancing, circuit dosage investigation, and novel selection schemes.

2.5 – Methods

2.5.1 – Media

Strains were propagated in LB broth and LB agar plates, with addition of 100 µg/ml ampicillin sodium salt, 50 µg/ml spectinomycin dihydrochloride pentahydrate, 25 µg/ml kanamycin sulphate, and/or 10 µg/ml trimethoprim if appropriate.
2.5.2 – Plasmids

Plasmids were constructed using BglBrick standard assembly[43]. Plasmid and strain descriptions are available in Chapter 6, and complete sequences are available through the Registry of Standard Biological Parts (http://partsregistry.org) [44]. The replicon of ColE2-P9[29] is referred to as ColE2 for convenience. The gamma origin of R6K is similarly referred to simply as the R6K origin.

2.5.3 – Genomic RBS Library Construction

Template plasmids BBa_J72204-BBa_J72184 and BBa_J72111-BBa_J72186, illustrated schematically in Figure 2.5, were first constructed. Splicing by overlap extension SOEing PCR [38] with degenerate oligos (Table 2.2) was used to generate RBS variants (NNNGGAANNNNNNRTG for pir and NNNAGAANNNNNNRTG for repA) of the cassettes on the template plasmids. The final PCR products were gel purified using Zymo columns according to the manufacturer’s instructions, and then used to modify the genome of strain MC1061 [45] by the procedure of Datsenko and Wanner [46]. The resulting libraries consisted of 10,000 members each, and were pooled before preliminary transformation with fluorescent protein expressing plasmids and analysis of fluorescence levels. Variants ultimately selected for full characterization were P1 transduced into MC1061 cells before further analysis to eliminate the initial fluorescent plasmid.

Figure 2.5 – Schematic of templates constructed for genomic integration. Plasmids BBa_J72204-BBa_J72184 and BBa_J72111-BBa_J72186 both contain: 1) a ~500 bp homology arm (HA) matching the 5' end of the genomic insertion site 2) a constitutive promoter (P<sub>CON</sub>) 3) the ORF of the trans-acting replication factor (Trans) 4) a terminator 5) a kanamycin resistance cassette (KnR) flanked by FRT sites and 6) a ~500 bp HA matching the 3’ end of the genomic insertion site.

Table 2.2 – Oligos used for library construction

<table>
<thead>
<tr>
<th>Oligo</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pir</td>
<td></td>
</tr>
<tr>
<td>Outer F</td>
<td>CACATGGCGACCAGATCAATAC</td>
</tr>
<tr>
<td>Outer R</td>
<td>ATTGCCGCAGGTGGAAC</td>
</tr>
<tr>
<td>Inner F</td>
<td>GGTACAGTGCTAGCGGATCTNNNGGAANNNNNNRTGAGCTCAAGGTCATGATGG</td>
</tr>
<tr>
<td>Inner R</td>
<td>GATCCGCTAGCCTACGTACC</td>
</tr>
<tr>
<td>repA</td>
<td></td>
</tr>
<tr>
<td>Outer F</td>
<td>GGATTTTCCTTGTTTCCAGAG</td>
</tr>
<tr>
<td>Outer R</td>
<td>GCTTACGGCTTTATATCGG</td>
</tr>
<tr>
<td>Inner F</td>
<td>CTGGTCAGTAAACAGCCCTNNAGAANNNNNNRTGAGCGCGGTACTTC</td>
</tr>
<tr>
<td>Inner R</td>
<td>AGGGCTCGTTACTGACGAG</td>
</tr>
</tbody>
</table>
2.5.4 – Plasmid Copy Number Estimation by qPCR

Plasmid copy number per genome equivalent was estimated using the relative quantitation method described previously [47]. Briefly, cells were subcultured 1:100 into fresh media and grown until mid-log or stationary phase before total DNA isolation using QIAamp DNA Mini kits (Qiagen) according to the manufacturer’s instructions. DNA samples and 10-fold serial dilutions of a purified calibrator plasmid bearing a single copy of both bla and ddx (BBa_J72109-BBa_J72187) were then amplified on an iCycler with iQ5 real-time PCR detection system (Biorad) using previously validated primer pairs [47] for both bla and ddx (bla F: 5’-CTACGATACGGGAGGGCTTAE3’ blaR: 5’-ATAAATCTGGAGCCGGTGAG-3’ ddxF: 5’-CGAGAAACTGCGATCCTTTA-3’ ddxR: 5’-CTTCATCAAGCGGTGTTCA-3’). Each reaction contained 25 µl: 12.5 µl Absolute QPCR SYBR Green Fluorescein Mix (Thermo Scientific), 1.25 µl each primer (10 µM), 3.75 µl H2O, and 5 µl sample DNA. Reaction conditions were as follows: an initial denaturation at 95°C for 10 minutes, followed by 40 cycles of 95°C for 10 seconds, 63°C for 15 seconds, and 72°C for 15 seconds. Measurements were taken at the end of each extension step. Copy numbers were calculated by using the calibrator standard curves to determine the quantity of plasmid (bla) and genome (dxs) DNA for a given sample in arbitrary units, and then calculating their ratio.

2.5.5 – Flow Cytometry

Cells grown to stationary phage were subcultured 1:100 in fresh media, grown until mid-log, resuspended in PBS, and then examined on a Coulter Epics XI-MCl instrument with a 488 nm excitation wavelength and 525 nm emission bandpass filter.

2.5.6 – Stability Analysis

Single colonies were picked and grown to stationary phase under selection. Cells were then subcultured 1:10^6 and grown back to stationary phase without selection, which corresponds to 20 generations of growth. The dilution and regrowth was repeated serially for 100 generations, at which point dilutions of cells were plated on non-selective media, and colonies were examined for sfGFP fluorescence as an indicator of plasmid presence.

2.6 – Acknowledgements

The authors would like to thank Dr. Tateo Itoh for providing the ColE2-P9 replicon. This work was supported by the National Science Foundation Synthetic Biology Engineering Research Center (SynBERC). JTK was supported by a National Science Foundation Graduate Research Fellowship.
Chapter 3 – Scalable Plasmid Transfer using Engineered P1-based Phagemids

Authors:

Joshua T Kittleson
Department of Bioengineering, University of California, Berkeley CA, United States

Will DeLoache
Department of Bioengineering, University of California, Berkeley CA, United States

Hsiao-Ying Cheng
Department of Bioengineering, University of California, Berkeley CA, United States

J. Christopher Anderson (corresponding author: jcanderson@berkeley.edu)
Department of Bioengineering, University of California, Berkeley CA, United States
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley CA, United States


3.1 – Abstract

Dramatic improvements to computational, robotic, and biological tools have enabled genetic engineers to conduct increasingly sophisticated experiments. Further development of biological tools offers a route to bypass complex or expensive mechanical operations, thereby reducing the time and cost of highly parallelized experiments. Here, we engineer a system based on bacteriophage P1 to transfer DNA from one *E. coli* cell to another, bypassing the need for intermediate DNA isolation (e.g. minipreps). To initiate plasmid transfer, we refactored a native phage element into a DNA module capable of heterologously inducing phage lysis. After incorporating known *cis*-acting elements, we identified a novel *cis*-acting element that further improves transduction efficiency, exemplifying the ability of synthetic systems to offer insight into native ones. The system transfers DNAs up to 25 kilobases, the maximum assayed size, and operates well at microliter volumes, enabling manipulation of most routinely used DNAs. The system’s large DNA capacity and physical coupling of phage particles to phagemid DNA suggest applicability to biosynthetic pathway evolution, functional proteomics, and, ultimately, diverse molecular biology operations including DNA fabrication.

3.2 – Introduction

Biological discovery and design relies on a wide array of software, hardware, and bioware tools. The last describes biological entities or their derivatives, such as DNA
manipulation enzymes, recombinant cell lines, and general transducing phages. Importantly, bioware such as yeast two-hybrid systems (Y2H) [48] reduce otherwise complex and expensive operations into simple, highly parallelized procedures. Absent bioware such as Y2H, identifying interactions between a target protein and ~3000 other proteins requires time- and resource-consuming expression and purification of each protein, followed by an appropriate in vitro binding assay. By contrast, traditional Y2H requires only plasmid transformation, population selection, and interaction pair sequencing. With the advent of next generation sequencing, Y2H has been extended to provide data on an entire interactome from a single experiment [44]. Similarly, the mating-assisted genetically integrated cloning (MAGIC) strategy takes advantage of bacterial conjugation, recombination, and selection to simplify the cloning process [49]. Incorporation of bioware into MAGIC eliminated the costs and complications associated with lysing cells, purifying DNA from the lysate, and prepping cells for DNA transformation, thereby achieving greatly improved throughput. These examples highlight how transformative jumps in throughput can be accomplished by performing otherwise slow, laborious, or expensive unit operations with appropriately designed bioware. Development of bioware tools compatible with high-throughput robotics for routine operations could dramatically reduce the costs and improve the throughput of diverse experimental approaches.

In particular, routine methods of DNA manipulation rely principally on *E. coli* for plasmid amplification and propagation, taking advantage of its fast growth, high competency, and high DNA yields. While generally effective, such methods are still subject to a number of constraints: isolation of high quality DNA from *E. coli* usually requires either centrifugation or vacuum application, analysis of the DNA frequently relies on gel electrophoresis, growth cycles of *E. coli* in liquid and solid media require on the order of 12 hours, and clonal selection methods require isolation of a single colony from a plate. These difficult hardware operations limit automatability, with expensive, specialized robotics required to achieve even moderate throughput using a 96-well format (e.g. the Qiagen BioRobot 8000). If the operations were instead implemented as bioware designed to function using purely liquid handling manipulations, they could instead be conducted at the nanoliter scale with throughput in the hundreds of thousands using technologies such as acoustic liquid handlers and microfluidics [50]. To overcome the inherent limitations of relying on bacteria alone, here we investigate using phage-based DNA transfer to address the essential problem of moving DNAs efficiently between cells.

Many phage can infect, reproduce, and lyse their hosts in less than an hour [51], and they have evolved to transmit genetic information from one host to another through aqueous intermediates. We selected phage P1 to use as a starting point to take advantage of its extensive characterization [52] and ability to package large DNAs from a well-defined packaging site. Because P1 employs a headful packing mechanism, any sized DNA up to 90 kb can be packaged into a P1 particle, since smaller DNAs can either be concatenated before uptake or packaged serially [53]. Further, another group has already demonstrated that a P1 phagemid, a plasmid containing *cis* phage elements, can be successfully packaged and delivered into new cells [54]. In that study, a temperature inducible P1 lysogen was used to generate particles containing a plasmid with a packaging site and a P1 lytic origin of replication. In this work, we employ a transcriptionally activated mechanism for inducing lysis to generate a small molecule responsive phagemid system. In an effort to improve efficiency, we also isolate an additional *cis* element that enhances phagemid transduction. Finally, we characterize behaviors of the system relevant to its potential applications. The resulting DNA transfer system operates at low volumes under isothermal conditions, and should find application in improving continuous selection schemes,
identifying protein-protein interactions, and streamlining diverse molecular biology operations including DNA fabrication.

**Figure 3.1 – Genetic interaction of phage lysogen and phagemid.** The interaction between a complete phagemid, J72103, and a simplified representation of the P1 genome is shown. For simplicity, only interactions relevant to the system behavior or mentioned in the main text are presented here. C1 is the master repressor of lysis, and can be inhibited by Coi. It also downregulates expression of the native coi, but not the refactored version on the phagemid. The phage lysogen provides both repL and pacA, which act on sites within their respective coding regions. The phagemid also expresses pacA, but has a truncated version of repL encoding only the cis element.

3.3 – Results and Discussion

3.3.1 – Transcription of coi Triggers the P1 Lytic Cycle

We sought a trans-acting protein capable of inducing the P1 lytic cycle, reasoning that transcription of such a protein would provide a mechanism for controlling phage induction. The P1 lysis/lysogeny decision is regulated by the activity a master repressor, C1. Although a complex genetic circuit controls C1 synthesis and activity during P1 infection [52], heterologous expression of a C1 inhibitory protein, Coi, blocks C1 activity, thereby inducing the lytic cycle [55]. Here, we similarly employ overexpression of coi from a phagemid to induce the P1 lytic cycle, and the interactions between a P1 lysogen and a coi-based phagemid are summarized in Figure 3.1. To validate this approach, we first refactored coi into a synthetic module activated by transcription, removing all sequence 5’ of the ribosome binding site, taken to be the 17 base pairs
immediately 5’ of the start codon [56], and adding a terminator 3’ of the stop codon. We then constructed plasmids with the refactored coi under the transcriptional control of one of two small molecule inducible promoters, P_BAD or P_RHA, which are induced by arabinose or rhamnose, respectively. Addition of the appropriate inducer to P1 lysogens transformed with these plasmids resulted in cell lysis, while uninduced controls and controls lacking a phagemid failed to lyse (Figure 3.2a). Consistent with previous results [55], this confirms that induced transcription of coi from phagemids initiates lysis.

**Figure 3.2 – Expression of coi induces lysis.** (a) MC1061 P1 lysogens harboring a plasmid with an arabinose-inducible coi (P_BAD-coi, J72110- J72094), a plasmid with a rhamnose inducible coi (P_RHA-coi, J72110- J72097), or no plasmid were incubated in the presence or absence of inducer and the OD_{600} monitored. Each construct was run in triplicate; error bars have been omitted for clarity, with all standard deviations falling below 16% of the mean, except the P_BAD-coi construct induced with arabinose, which had standard deviations as high as 67% of the mean. (b) P1 lysogenized MC1061 containing either a plasmid with only an arabinose-inducible coi (P_BAD-coi, J72110- J72094) or a complete phagemid (P_BAD-coi + cin + repL + pacA, J72110-J72103) were induced to lyse by addition of arabinose. The corresponding arabinose promoter activity, reported in relative promoter units (RPU), is plotted against the resulting percent lysis, calculated as 1 – (OD_{600} sample / OD_{600} uninduced control). Error bars indicate the standard deviation of quadruplicate samples.

Although we focus here on the use of small-molecule inducible promoters to control lysis, other applications may require the use of a different triggering mechanism. To aid in the development of alternative coi-based lysis circuits, we quantified lysis induction efficiency as a function of P_BAD transcriptional activity. As illustrated in Figure 3.2b, lysis of both an arabinose-driven coi-only phagemid and a phagemid bearing additional cis elements (coi + cin + repL +
*pacA*, discussed below) exhibited dose dependence, appearing to saturate at the maximal observed promoter strength of 0.4 relative promoter units (RPUs) [57]. We expect that any alternative phagemid design that provides the same level *coi* transcription would also function to induce lysis.

**Figure 3.3 – Impact of *cis* elements on packaging.** Test phagemids harboring various combinations of putative *cis* packaging elements were competed against a reference phagemid containing all of the elements (J72109-J72105: *cin* + *repL* + *pacA*). The resulting ratio of test phagemid to reference phagemid indicates the relative efficiency of test phagemid packaging. Error bars indicate the standard deviation of triplicate biological samples. Starred pairs show a statistically significant difference (p<0.05 in an unpaired two-tailed t-test). The test phagemids are J72113-J72106: *coi* + *rfp*, J72113-J72094: *coi*, J72113-J72102: *coi* + *cin*, J72113-J72095: *coi* + *repL*, J72113-J72115: *coi* + *cin* + *repL*, J72113-J72101: *coi* + *pacA*, J72113-J72116: *coi* + *cin* + *pacA*, J72113-J72096: *coi* + *pacA* + *repL*, and J72113-J72103: *coi* + *cin* + *repL* + *pacA*.

### 3.3.2 – *cin* is a *cis* Element that Improves Phagemid Transduction

Motivated by the observation that plasmids not harboring phage *cis* elements are also transduced, albeit infrequently, we sought to improve the relative efficiency with which desired plasmid DNA is transduced. Incorporation of a packaging site (which resides within *pacA*) and a lytic origin of replication (which resides within *repL*) both enhance packaging [54]. Although no other packaging or delivery enhancers have been reported, we reasoned that there might be additional regions of the P1 genome that improve DNA packaging or delivery. We therefore inserted a library of P1 genomic fragments into a phagemid harboring complete *pacA* and *repL* ORFs. The library was enriched for elements that improve transduction by generating phagemid
lysate from pooled library members, and then infecting naïve lysogens with the lysate. After several rounds of such enrichment, individual clones were isolated and sequenced to identify the inserted genomic fragment. Sequences of interest were PCR-amplified out of the P1 genome and cloned into phagemids, which were subsequently tested individually for improved function. One region, which contained the cin ORF, consistently showed improvement (data not shown). Although there is no published cis role for cin in the packaging or delivery of P1 phage DNA, it has been shown to encode a site-specific recombinase responsible for switching the host specificity of the phage [58]. To determine if the improvement was an unexpected function of the coding sequence or an overlying cis element, a noncoding version of the ORF lacking a start codon and bearing a nonsense mutation was created. Inclusion of the noncoding cin DNA improved the relative transduction efficiency of phagemids harboring coi and any combination of repL and pacA by 3-11 fold (Figure 3.3), suggesting that cin operates as a cis element. We also examined the possibility that increased plasmid size alone might lead to improved transduction, but inclusion of a gene coding for red fluorescent protein had no effect on transduction efficiency. Importantly, a phagemid harboring all three cis elements showed an approximately 1600 fold improved yield over a plasmid lacking any cis elements. This suggests that the system very specifically transduces desired phagemid DNA.

Figure 3.4 – Impact of volume on phagemid efficiency. (a) MC1061 P1 lysogens containing a complete phagemid (J72110-J72103) were induced to lyse at 2000 µL, 200 µL, 50 µL, and 10 µL scale in 24 well blocks, 96 well plates, 384 well plates, and 1536 well plates, respectively, and the lysate used to transduce naïve JTK029 cells for titer determination. Error bars indicate the standard deviation of biological triplicates. Differences between conditions are not statistically significant (p>0.05 in an unpaired two-tailed t-test). (b) P1 lysogenized JTK160C harboring plasmid J72111-J72098 was transduced with two-fold serial dilutions of J72110-J72103 lysate from (a) in the presence of two-fold serial dilutions of arabinose (highest concentration, 13 mM) in a 384 well plate. Tetrazolium violet was added to visualize unlysed cells.

3.3.3 – Phagemids Exhibit Robust and Faithful Transduction
Because we envision utilization of the phagemid in high-throughput liquid handling operations with diverse genetic material, it is important for phagemid transduction to be robust to both reaction volume and phagemid size. To establish the effect of reaction volume on lysis, we generated lysate from P1 lysogenized cells bearing phagemids in different vessels, and used the lysate to transduce naïve cells to titer viable phagemid-carrying phage particles (Figure 3.4a). Although the average titer dropped approximately 2-fold between a 2000 µL lysis volume and a 200 µL lysis volume, this difference is not statistically significant (p>0.05), and there is little impact on the average titer from further scale reduction. To ascertain if transduction also operates at small volumes, we tested the ability of phagemid lysate to transduce naïve lysogens and initiate lysis in a 384 well plate (50 µL scale). Figure 3.4b illustrates that a high titer of phagemid, in the presence of sufficient inducer, leads to complete lysis of the recipient cells. This suggests that transduction functions at small volumes, as expected.

To probe the size constraints of the system, we tested the transduction efficiency of a modest (10kb), a medium (14kb), and a large (25kb) phagemid; larger constructs are not readily supported by the p15a origin of replication [26]. We observed that although the smallest phagemid transduced with 2.5 fold higher efficiency than the larger phagemids, all of the phagemids generated hundreds to thousands of transductants (Figure 3.5). Therefore, this system is capable of operating on large DNAs, such as those encoding multi-gene genetic circuits.

**Figure 3.5 – Impact of phagemid size on efficiency.** MC1061 P1 lysogens containing one of three phagemids (J72114-J72100 – 9729 bp, J72114-J72090 – 13938 bp, or J72114-J72104 – 25114 bp) were induced to lyse with arabinose, and the lysate used to transduce naïve MC1061 cells for titer determination. Error bars indicate the standard deviation of biological triplicates. The 9729 bp phagemid is significantly different from the other two phagemids (p<0.05 in an unpaired two-tailed t test), which are not significantly different from each other (p>0.05 in an unpaired two-tailed t test).

Some applications of the phagemid system may involve manipulation of libraries of genetic constructs. Under conditions such as those in Figure 3.4a, 1 µL of phagemid lysate is sufficient to generate up to 10,000 transductants. Scaled up to higher volumes, this implies that 100 ml of lysate could generate up to $10^9$ clones. However, library manipulations should also preserve the relative abundance of each clone. As a simple probe of this functionality, we generated a library of phagemids expressing different levels of green fluorescent protein (GFP) in P1-lysogenized cells, illustrated in Figure 3.6a. We pooled the clones, induced lysis, and
transduced naïve P1-lysogenized cells with the lysate. The resulting clones (Figure 3.6b) show a similar diversity of GFP expression levels. To better quantify the fidelity of library transfer, we used a mixture of GFP-expressing and non GFP-expressing phagemids to generate a library with 10% GFP-expressing clones. Lysate generated from pooled library members was used to transduce naïve P1-lysogenized cells, and resulted in an average of 9% GFP-expressing clones, which is not significantly different from the original library (p>0.05, Figure 3.6c). The ability to produce large numbers of clones without significantly perturbing the ratio of variants confirms the utility of phagemids for transduction of genetic libraries.

**Figure 3.6 – Library transfer using the phagemid.** MC1061 P1 lysogens containing a library of GFP expressing phagemids (J72110-J72152 library) (a) were pooled, induced to lyse with arabinose, and the lysate used to transduce naïve MC1061 P1 lysogens (b). (c) MC1061 P1 lysogens were transformed with a mixture of GFP expressing (J72113-J72152) and non-GFP expressing (J72113-J72103) phagemids, and the percent of GFP expressing clones counted (the original library). The clones were pooled, induced to lyse, and the lysate used to transduce naïve MC1061 P1 lysogens. The resulting clones (the transduced library) were counted, and the percent expressing GFP calculated. Error bars indicate the standard deviation from three replicates. The difference between the original library and the transduced library is not statistically significant (p>0.05 in an unpaired two-tailed t test).

**3.4 – Summary and conclusions**

We have engineered and characterized a P1-based phagemid that requires only liquid handling for the isothermal, high efficiency transfer of plasmids from one cell to another. As part of this process, we isolated an additional *cis* element that lies inside of the *cin* gene and enhances transduction of the phagemid. Because there is no known role for *cin* in the packaging or delivery of P1 phage DNA, the biological mechanism of improved transduction remains to be elucidated, illustrating how engineering novel synthetic systems can probe our understanding of native systems.
By refactoring the natural transcriptional regulation of coi, we gain control of the lysogeny-to-lysis switch. Here, we demonstrated simple control by a heterologous input, arabinose or rhamnose, using an inducible promoter to drive coi expression. In theory, however, any genetic circuit could be used. For example, recent work by Esvelt and colleagues [59] demonstrated the power of connecting phage viability to an encoded gene function with the development of phage-assisted continuous evolution (PACE). An analogous approach could be used to evolve an entire biosynthetic pathway encoded on a P1 phagemid by using a biosensor [60] for the product of the pathway to trigger expression of coi.

Addition of cin to the phagemid improved transduction efficiency, enabling a phagemid bearing all relevant cis elements to be transduced 1600-fold more efficiently than a plasmid lacking any such elements. Although further work is required to elucidate the relative packaging efficiency of phagemids and wild-type P1 phage genomes, which may be important to continuous selection schemes such as PACE, the current system is clearly sufficient to transfer both individual clones and libraries of genetic elements. The ability to specifically couple at least 25 kb of DNA to a phage particle takes a step toward enabling P1 phage display for functional proteomics. Based on work with other well characterized phages [56, 61-63], we speculate that phagemid-encoded protein complexes could be attached to the particle surface, e.g. via a short tag [64], allowing functional complexes to be enriched by targeted binding to individual complex components.

Transcriptional control of lysis provides a final advantage: isothermal phage induction, which confers compatibility with a simple liquid handling robotics including microfluidic technologies. The robustness of the system at microliter volumes suggests the potential for operation on nanoliter scale platforms. DNA transfer, in combination with as yet undeveloped bioware operations such as restriction and ligation, may ultimately provide a mechanism for fast, cheap DNA fabrication. Although development of such bioware is challenging to engineer today due to the general difficulties of creating complex genetic systems [2], improvements in our ability to design [65] and fabricate [10] genetic systems will continue to make bioware development easier. An initial investment in the development of additional liquid handling-only bioware tools for common procedures would reduce the cost of genetic engineering to the purchase of a standard liquid handling platform, accelerating development of useful biotechnologies.

3.5 – Methods

3.5.1 – Plasmids, Strains, Phage, and Growth Media

Plasmids used in this study are described in Chapter 6, and were constructed using BglBrick standard assembly [43]. E. coli strain MC1061 [45] or derivatives were used for all studies. Derivatives were generated by the procedure of Datsenko and Wanner [62]. Full descriptions of plasmids and strain modifications are available in Chapter 6, and DNA sequences are available at the Registry of Standard Biological Parts (http://partsregistry.org) [44]. For brevity, the BBa_ prefix of sequence names has been omitted throughout the manuscript. P1kc [66] (obtained in strain KL739 from the Yale Coli Genetic Stock Center) was used for experiments involving phage lysogens, and will be referred to simply as P1 for convenience. P1vir [67] (gift from Carlos Bustamante) DNA was used for library preparation. Bacterial cells
were grown in Luria-Bertani medium (LB) or phage lysis medium (PLM; LB containing 100 mM MgCl₂ and 5 mM CaCl₂).

3.5.2 – Generation of Phage Lysates

Stationary phase cultures of a P1 lysogenized strain harboring a phagemid were diluted 100-fold into PLM and incubated for 1 hour at 37°C. Lysis was then induced by addition of either 13 mM arabinose or 10 mM rhamnose. After 1-4 hours of further incubation at 37°C, chloroform (2.5% v/v) was added and the culture was vortexed for 30 seconds. The culture was then clarified by centrifugation (12,100 g for 90 seconds), and the supernatant recovered. For characterization of lysis efficiency in different volumes, the chloroform and clarification steps were omitted, and the crude lysate was instead used directly for subsequent transduction.

3.5.3 – Transduction of Phage Lysates

A stationary phase culture of the host strain was isolated by centrifugation and resuspended to an OD₆₀₀ of 2 in PLM. An equal volume of phage lysate was mixed with resuspended cells and allowed to adsorb at 37°C without agitation for 30 minutes. The reaction was quenched with 5 volumes of 2YT medium containing 200 mM sodium citrate, pH 5.5, and the infected cells were incubated for an additional hour at 37°C before titering transductants on plates with appropriate antibiotics.

3.5.4 – Timecourse of Phage Lysis

Stationary phase cultures of MC1061 P1 lysogens were diluted 100-fold into PLM. After growth for 1 hour at 37°C with agitation, 200 µL of cells were transferred to a Corning flat-bottom 96 well microplate and induced by addition of either 13 mM arabinose or 10 mM rhamnose. Culture OD₆₀₀ was then monitored every 5 minutes in a Tecan Sapphire instrument with continued incubation at 37°C with agitation. OD₆₀₀ measurements were then normalized to the first OD₆₀₀ reading (OD₆₀₀ at time t divided by OD₆₀₀ at time 0).

3.5.5 – Measurement of Relative Promoter Units (RPU)

The RPU of the Pₐ₅₆₉₉ promoter with different levels of arabinose induction (0-13 mM) was measured essentially as previously described [57]. Briefly, samples of stationary phase MC1061 P1 lysogen cells harboring a p15a plasmid with either a constitutive reference promoter (Pₐ₅₆₉₉, J72110-J72107) or Pₐ₅₆₉₉ (J72110-J72108) driving expression of green fluorescent protein were diluted 100-fold into fresh inducing media and grown until mid exponential phase. At that point, the OD₆₀₀ (OD) and fluorescence (F) were measured. RPU was then calculated using equation 10 from [57], with the approximation that all relevant growth rates are equal: \( \frac{F(P_{BAD})/OD(P_{BAD})}{F(P_{REF})/OD(P_{REF})} \).

3.5.6 – Visualization of Serial Lysis

Stationary phase culture of the host strain was diluted 100 fold into PLM and incubated for 75 minutes at 37°C. Subsequently, 25 µL aliquots of cells were arrayed into a Genetix flat-
bottom 384 well microplate and incubated with 25 µL of phagemid lysate for 2 hours at 37°C in the presence of arabinose. Cells were then stained by addition of 0.04% w/v tetrazolium violet, and the plate incubated for an additional 2 hours at 37°C before image acquisition.

3.5.7 – Phagemid Library Construction and Enrichment

To isolate P1 vir DNA, 250 µL P1 vir lysate was treated with 5 µL of Qiagen buffer L1 (300 mM NaCl; 100mM Tris-Cl, pH 7.5; 10 mM EDTA; 0.2 mg/ml BSA; 20 mg/ml RNase A; 6 mg/ml DNase I) and incubated for 30 minutes at 37°C. After incubation, 400 µL 4% SDS was added, and the mixture further incubated for 15 minutes at 55°C. 350 µL of Qiagen buffer N3 was then added, and the mixture applied to a Qiagen Qiaprep Spin Column. DNA isolation then proceeded according to the manufacturer’s instructions for plasmid DNA purification.

Purified P1 vir DNA was then amplified using Phi29 polymerase. A typical reaction consisted of 2 µL 10x Phi29 buffer, 5 U Phi29 polymerase, 400 ng thiophosphate protected random hexamers [68], 500 µM each dNTP, approximately 25 ng template DNA, and water to 20 µL. The reaction mixture was incubated at 30°C for 16 hours, and then purified using Zymo-Spin I columns according the manufacturer’s instructions.

Amplified P1 vir DNA was then partially digested with Sau3AI, and fragments of 1-4 kilobases ligated into BamHI digested and CIP treated J72112-J72099 plasmid DNA. Plasmids were inserted into a JTK030 P1kc lysogen by transformation, generating a library of 3 x 10^6 members. Library members were pooled and then used to generate phagemid lysate. This lysate was used to transduce phagemids into naïve JTK030 P1kc lysogen cells, and the cycle of pooling, lysis, and transduction repeated.

3.5.8 – Competition between phagemids

MC1061 P1 lysogens were co-transformed with two phagemids, a non-inducible “reference” phagemid (J72109-J72105) composed of cin, pacA, and repL, and a “test” phagemid (in vector J72113) composed of arabinose-inducible coi and a subset of the elements pacA, repL, and cin. Phage lysate was generated as described above, and then used to transduce naïve MC1061 cells for titering of each phagemid.

3.5.9 – Construction of library of GFP expressing phagemids

Splicing by overlap extension SOEing PCR [38] with degenerate oligos (Outer forward: 5’-GTATCAGGAGCAGAATTTCCAG-3’, Outer reverse: 5’-ATTACCCTTTCGAGTGCACG-3’, Inner forward: 5’-GCACGGATCTTAGCTACTAGAAGANNNNGAANNNNNNTGCGTAAAGGCGAAGCGC-3’, Inner reverse: 5’-TCTTTTCCTAGTAGCTAAGATCGTGGTGC-3’) on template plasmid J72111-J72155 was used to generate RBS variants (NNNGAANNNNNNTG) of the GFP expression cassette. The SOEing PCR product was digested with EcoRI and BamHI and cloned into the EcoRI and BamHI sites of vector J72154. The resulting library of clones was pooled before DNA extraction and subsequent assembly with plasmid J72153-J72103 using the 2ab strategy [69] to generate a library of variants of J72110-J72152.
3.6 – Acknowledgements

The authors would like to thank Caroline Westwater for thermoinducible P1 phage (P1Cm cIts.100), and Carlos Bustamante for P1vir. This work was supported by the National Science Foundation Synthetic Biology Engineering Research Center (SynBERC). JTK received support from a National Science Foundation Graduate Research Fellowship and a Siebel Scholar award. WD received support through the Amgen Scholars Program.
Chapter 4 – Toward Engineering *E. coli* as a Cancer Therapeutic

4.1 – Abstract

The ability of bacteria to specifically target tumor tissue and intimately interact with host cells has led to development of several bacterial cancer therapies effective in animal models. However, human clinical trials suggest that further bacterial engineering is required to create a clinically viable bacterial therapeutic. In this study, we work toward engineering a laboratory strain of *E. coli* that both exhibits an extended lifetime after systemic administration in a mouse model and interacts with and kills tumor cells. First, we attempt to engineer a strain to express both an O-antigen and a K-capsule to protect against attack by the immune system. Although we both confirm that capsular polysaccharides are essential for extended survival and transfer functional O-antigens into a laboratory strain, transfer of a protective K-capsule remains technically challenging. Second, we demonstrate that a strain engineered to invade tumor cells and release a ribonuclease is capable of killing cultured tumor cells *in vitro*. However, the strain failed to affect tumor growth in a mouse model, and further development will be required to identify and overcome shortcomings in the current design.

4.2 – Introduction

Although several compounds are effective at killing cancer cells, their therapeutic potential is limited by off-target activity and inadequate distribution to tumor tissue [70]. Recognizing these limitations, several next-generation drug delivery strategies have been developed to more precisely deliver payload molecules to target cells, thereby avoiding toxic side effects. For example, complex branched molecules called dendrimers can be designed to interact with specific cell types and release a payload only under specific conditions [71]. An alternative strategy employs liposomes, which enclose a drug payload inside of a lipid membrane. The membrane can then be modified by incorporation of targeting elements to ensure more specific delivery to tumor cells [72]. While promising, these targeting approaches face important challenges, including expensive synthesis, reliance on passive distribution, and a limited set of functions that can be statically encoded in the structure of the delivery vehicle. A novel approach to overcome these limitations is the use of bacteria as a therapeutic delivery vehicle. Because they can dynamically respond to a variety of physiologically relevant stimuli, self-replicate, and intimately interact with host cells and systems, bacteria offer a powerful platform for therapeutic development.

Bacteria are particularly well suited to development as therapeutics for cancer. A diverse group of both anaerobes and facultative anaerobes selectively localize to solid tumors after systemic administration [73-76]. For facultative anaerobes, this likely stems from an enhanced delivery and retention effect, coupled with a lack of immune surveillance in tumors [77]. Researchers have taken advantage of this natural localization ability to engineering bacteria as tumor visualizing agents. For example, *E. coli* expressing luciferase genes have been used to identify small metastatic nodules in mice [76]. Beyond localization and imaging, bacteria are known to be capable of causing tumor regression. In studies in the late 1800’s [78], it was observed that some patients bearing tumors, upon being deliberately infected with bacteria, would experience tumor regression. However, the treatment had considerable negative side
effects and only applied to a narrow range of tumor types, and was therefore abandoned in favor of chemotherapy.

Recent renewed interest in exploring unconventional cancer therapeutics has led to novel approaches for using bacteria as cancer therapeutics (reviewed in [77]). Non-pathogenic (or attenuated) bacteria have been used directly [79], in combination with conventional therapy [80], engineered to express a tumoricidal agent [81, 82], and engineered to express an enzyme that converts a prodrug into an active drug [83-86]. These approaches have yielded considerable improvements in mouse models, but failed to provide significant improvement in the clinic [77]. The case of attenuated variants of *Salmonella spp.* is particularly relevant: they have shown considerable effectiveness in mouse models, including complete regression of solid tumors [79] and elimination of hundreds of lung metastases [87]. However, in clinical trials [88, 89], these salmonella failed to robustly localize to patient tumors, and no effect on tumor growth was observed. One plausible explanation for this failure is that the immune system of humans is much more efficient at clearing the attenuated *Salmonella*, prevent distribution to tumor sites.

Failures in the clinic motivate us to engineer a strain to be more resistant to clearance by the immune system. Because of its safety and genetic tractability, we focus on engineering a non-pathogenic laboratory strain of *E. coli* to have a reduced immune clearance rate, and then confer it with a tumor destroying capability. In previous work [90], we developed *E. coli* capable of invading cancer cells *in vitro* and delivering benign reporter proteins. Here, we present progress on extending the *in vivo* survival of *E. coli* by incorporation of capsular polysaccharides. We then assess the ability of *E. coli* bearing a cytotoxic payload to destroy tumor cells both *in vitro* and *in vivo*.

### 4.3 – Results

#### 4.3.1 – Encapsulation of *E. coli*

*E. coli* causing extra-intestinal infections have two layers of polysaccharide coating that protect them from host defenses (reviewed in [91]). The O-antigen, an oligosaccharide extension of lipopolysaccharide, and the K-capsule, an additional external polysaccharide coat, function in concert to provide protection against several innate host immune responses, including both classical and alternative complement pathways and opsonophagocytosis. Strains made deficient in these protective coatings are consistently cleared from the host faster than their wild-type counterparts [92, 93].

There are numerous variants (serotypes) of both O-antigens and K-capsules, each with a unique biochemical composition. Of particular interest is the K1 capsule, a serotype overrepresented in patients suffering from meningitis [91] and causing high levels of bacteria in the bloodstream as part of its pathogenesis [94]. The K1 capsule is an effective defense mechanism because it is composed of repeating units of 2,8 linked sialic acid, a structure that both binds inhibitors of the complement cascade and mimics structures naturally found in the host [91]. Bacteria bearing K1 capsules are usually found with O-antigens of the type O1, O7, O16, or O18, though the reason for this association is unknown [95]. The genes clusters required for synthesis of all of these capsules are well-understood [96], making them excellent candidates for further manipulation.

Several lines of evidence suggest that constitutive expression of the protective layers is undesirable. While the polysaccharide coats sterically shield the bacteria from attack by the
immune system [97], it also hides underlying factors required for intimate interaction with host cells [98]. Consequently, capsule bearing bacteria often shed their protective layering during later stages of pathogenesis [99, 100]. Similarly, we expect that applications involving cell-cell adhesion or receptor mediated signaling will demand more intimate interactions than are permitted by the outer bacterial coats. Anticipating a need for dynamic control of capsule expression, we take an approach of placing the O-antigen and K-capsule under heterologous control by deleting a single gene from each gene cluster required for capsule synthesis, and complementing that gene on a plasmid.

Figure 4.1 – Strategies for manipulation of capsular polysaccharides. (a) To make a lab strain bearing K1 and O16, the neuS and wbbL genes were replaced with an antibiotic resistance marker in strains EV36 and BOS12, respectively. The genomic regions containing the marked biosynthetic pathways were then serially P1 transduced into MC1061. (b) An O7:K1 natural isolate was modified by deletion of either neuS or manB to make derivatives missing either the K1 capsule or O7 antigen, respectively. The neuS deletion strain was then used to make a double neuS manB deletion strain. In parallel, the genome of the O7:K1 isolate was used as a template to make plasmids bearing either the complete O7 or K1 synthesis cluster. Strains used in this study are listed in Chapter 6.

Because it remains unclear if encapsulation is sufficient or merely necessary to confer an extended survival phenotype, we approached capsule synthesis using both additive and subtractive strategies (Figure 4.1). First, organisms reported to express either a K1 capsule (EV36)[101] or O16 (Bos12)[102] were modified to replace either neuS or wbbL, respectively, with an antibiotic resistance marker. Transduction by bacteriophage P1 was then used to transfer both synthesis clusters into a lab strain of E. coli, MC1061. Capsule synthesis in this strain was controlled by addition of a plasmid expressing neuS, wbbL, or both. Second, plasmids containing complete synthesis clusters for O7 and K1 were constructed, using a natural O7:K1 isolate as a template. Finally, the natural O7:K1 isolate was modified by deletion of neuS, manB (required for O7 synthesis), or both. Capsule synthesis in these strains was similarly controlled by addition of a plasmid expressing one or both of the deleted genes.

To detect the presence of polysialic acid on the bacterial surface, strains were tested for sensitivity to bacteriophage K1E, which requires polysialic acid to adsorb to its host [103, 104]. To test for protection of the bacterial membrane by encapsulation, strains were assayed for their
ability to survive treatment with horse serum, bacteriophage T4, and/or bacteriophage T7 (table 1). As expected, strains bearing a complete biosynthetic pathway for either O16 or O7 show protection against membrane-mediated attacks, while unencapsulated strains are consistently killed. The only exception is expression of the O7 biosynthesis cluster in MC1061, which failed to confer protection. This can be readily explained by documented mutations in the galactose synthesis machinery of MC1061 [105], since galactose is a subunit of the O7 structure [106]. MG1655, a K12 strain with intact galactose synthesis machinery [107], shows protection with the O7 biosynthetic pathway. K1 capsule expression, however, proved more challenging. Although the K1 capsule transferred from EV36 into MC1061 led to K1E sensitivity, it did not confer protection to membrane attack. The biosynthetic cluster for K1 cloned onto a plasmid did not confer K1E sensitivity to MC1061 or MG1655. Because it only conferred partial K1E sensitivity to an O7:K1 △neuS knockout strain, we suspect that an undetected sequence error or complex regulation may account for a lack of activity in laboratory strains. Only the O7:K1 natural isolate, and derivatives thereof, exhibited both K1E sensitivity and protection from membrane attack in the absence of a functional O-antigen.

Although the in vitro assays suggested that the K1 capsule was only partially functional in laboratory bacteria, we proceeded to assess whether a functional O-antigen and partially functional K-capsule would provide an extended half-life in vivo. Survival of selected strains 120 minutes after systemic administration to a mouse is illustrated in Figure 4.2. The results are largely consistent with the in vitro assays: less than 0.05% of MC1061, either unaltered or with O16 and K1 capsules, survived, while up to 72% of derivatives of the O7:K1 survived. Deletion of a gene required for either K1 capsule or O7 antigen synthesis from the O7:K1 isolate led to ~100 fold reduced survival, while complementation of the deleted gene restored a high survival rate, as expected. Further K-capsule manipulation is required to improve the survival of a lab strain.

Figure 4.2 – Survival of encapsulated bacteria in a mouse. Strains were systemically injected into mice, and the mouse blood sampled 120 minutes later for cfu. Error bars are the standard error of the mean for at least 3 replicates. Strains and plasmids used, from left to right, are: MC1061; MC828E + pSB1A2-BBa_J72199; ATCC 23503; JTK041; JTK011; JTK041; JTK035 + BBa_J72109- BBa_J72193; JTK011 + pSB1A2-BBa_J72190; JTK041 + BBa_J72109- BBa_J72192.
Table 1 – *In vitro* assays of capsular polysaccharides

<table>
<thead>
<tr>
<th>Strain</th>
<th>Strain Description</th>
<th>Plasmid</th>
<th>Plasmid description</th>
<th>K1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>S/T4/T7&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC1061</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>EV36</td>
<td>K1</td>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Bos12</td>
<td>O16:K92</td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MC828Q</td>
<td>MC1061 O16(ΔwbbL) K1(ΔneuS) ΔmsbB Δfim ΔtonB</td>
<td>BBa_J72208-BBa_J72190</td>
<td>neuS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MC828Q</td>
<td>MC1061 O16(ΔwbbL) K1(ΔneuS) ΔmsbB Δfim ΔtonB</td>
<td>pSB1A2-BBa_J72191</td>
<td>wbbL</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MC828Q</td>
<td>MC1061 O16(ΔwbbL) K1(ΔneuS) ΔmsbB Δfim ΔtonB</td>
<td>BBa_J72208-1716023</td>
<td>wbbL, neuS</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 23503</td>
<td>O7:K1</td>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>JTK011</td>
<td>ATCC 23503 ΔneuS</td>
<td></td>
<td>pSB1A2-BBa_J72190</td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>JTK035</td>
<td>ATCC 23503 ΔmanB</td>
<td></td>
<td></td>
<td>nd</td>
<td>+</td>
</tr>
<tr>
<td>JTK037</td>
<td>ATCC 23503 ΔmanB ΔneuS kanR</td>
<td></td>
<td></td>
<td>nd</td>
<td>-</td>
</tr>
<tr>
<td>JTK041</td>
<td>ATCC 23503 ΔmanB ΔneuS</td>
<td></td>
<td></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>JTK041</td>
<td>ATCC 23503 ΔmanB ΔneuS</td>
<td>BBa_J72109-BBa_J72192</td>
<td>manB, neuS</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>JTK050</td>
<td>ATCC 23503 ΔmanB ΔneuS trans&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>JTK050</td>
<td>ATCC 23503 ΔmanB ΔneuS trans</td>
<td>BBa_J72109-BBa_J72193</td>
<td>manB</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>JTK050</td>
<td>ATCC 23503 ΔmanB ΔneuS trans</td>
<td>BBa_J72109-BBa_J72194</td>
<td>neuS</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>JTK050</td>
<td>ATCC 23503 ΔmanB ΔneuS trans</td>
<td>BBa_J72109-BBa_J72192</td>
<td>manB, neuS</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>MG1655</td>
<td></td>
<td></td>
<td></td>
<td>nd</td>
<td>-</td>
</tr>
<tr>
<td>MG1655</td>
<td>BBa_J72110-BBa_J72169</td>
<td>O7</td>
<td></td>
<td>nd</td>
<td>+</td>
</tr>
<tr>
<td>MC1061</td>
<td>BBa_J72110-BBa_J72169</td>
<td>O7</td>
<td></td>
<td>nd</td>
<td>-</td>
</tr>
<tr>
<td>MC1061</td>
<td>BBa_J72209-BBa_J72170</td>
<td>K1</td>
<td></td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>MG1655</td>
<td>BBa_J72209-BBa_J72170</td>
<td>K1</td>
<td></td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>JTK050</td>
<td>ATCC 23503 ΔmanB ΔneuS trans</td>
<td>BBa_J72209-BBa_J72170</td>
<td>K1</td>
<td>+/-</td>
<td>nd</td>
</tr>
</tbody>
</table>

<sup>a</sup> + indicates bacterial survival, or lack of a K1 capsule, while - indicates bacterial killing, or presence of a K1 capsule. +/- indicates a partial killing phenotype. nd not determined.

<sup>b</sup> + indicates bacterial survival when challenged with one of: serum, T4 phage, or T7 phage. - indicates bacterial killing when challenged with one of: serum, T4 phage, or T7 phage. nd not determined.

<sup>c</sup> *trans* is a highly transformable phenotype.
4.3.2 – Localization to Tumors and Plasmid Maintenance

In mouse tumor models, unencapsulated lab strains of E. coli have been reported to localize to solid tumors [108]. However, plasmids relying on antibiotic selection are sometimes lost during bacterial growth in the tumor. To overcome this limitation, other groups have exploited E. coli’s requirement for diaminopimelic acid (DAP) [109, 110]. If the genomic DAP biosynthesis pathway is disrupted by deletion of a gene, the resulting E. coli require either exogenous addition of DAP or complementation of the deleted gene on a plasmid. Since DAP is not known to be produced in mammalian systems, this provides a powerful selective force in vivo. We therefore deleted dapD from the genome, and provided a complementing copy on the plasmid backbone. To validate bacterial localization to tumors and plasmid maintenance, laboratory E. coli bearing various plasmids were administered to mice bearing solid tumors. In mice where bacteria were administered systemically (i.v. injection), 42% of tumors showed intensive bacterial colonization, while in mice where bacteria were administered directly into tumors, 86% showed intensive colonization (table 2). Without the use of the dapD plasmid maintenance system, none of the bacterial populations maintained all of their plasmids. However, using the dapD system, all of the bacteria maintained the dapD carrying plasmid, as expected. These results confirm that our bacteria can successfully colonize tumors and maintain a desired plasmid.

Table 2 – Bacterial Tumor Colonization and Plasmid Maintenance

<table>
<thead>
<tr>
<th>Strain</th>
<th>Strain Description</th>
<th>Plasmids</th>
<th>Injection</th>
<th>tumors with bacteria</th>
<th>bacteria with all resistances</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC1061</td>
<td></td>
<td>BBa_J72109-</td>
<td>i.v.</td>
<td>2 of 6</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72210-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72176</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72113-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC1061</td>
<td></td>
<td>BBa_J72195</td>
<td>i.v.</td>
<td>2 of 5</td>
<td>0 of 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC1061</td>
<td></td>
<td>BBa_J72196</td>
<td>i.t.</td>
<td>8 of 11</td>
<td>n/a</td>
</tr>
<tr>
<td>JH5-18</td>
<td>MC1061 inv^a</td>
<td>BBa_J72109-</td>
<td>i.t.</td>
<td>4 of 5</td>
<td>0 of 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72113-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JTK174</td>
<td>MC1061 inv2 ΔdapD</td>
<td>BBa_J72211-</td>
<td>i.t.</td>
<td>6 of 6</td>
<td>6 of 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72198</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JTK176</td>
<td>MC1061 inv3 ΔdapD</td>
<td>BBa_J72211-</td>
<td>i.t.</td>
<td>6 of 6</td>
<td>6 of 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BBa_J72198</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a inv, inv2, and inv3 denote alleles of invasin
Figure 4.3 – *In vitro* killing of tumor cells. HeLa cells were exposed either to bacteria (JH4-C7) carrying the payload delivery plasmid (BBa_J72218-BBa_J72197) and a payload plasmid (nothing, Cytolysin: BBa_J72212-BBa_J72200, BamHI: BBa_J72109-BBa_J72201, Barnase: BBa_J72212-BBa_J72195), or to staurosporine. Cell killing, as evidenced by detachment of cells, was determined 24 hours later. Barnase was later tested against B16F10 and HCT116 cells and found to be equally effective.

4.3.3 – Engineered *E. coli* Kill Tumor Cells *In Vitro*, but not *In Vivo*

We have previously described development of a bacterial system for delivering proteins to mammalian cell *in vitro*[90]. Briefly, *E. coli* was engineered to express Invasin from *Yersinia pseudotuberculosis*, enabling uptake by cancer cells overexpressing β integrin. Upon entry into an endosome, the bacteria express Holin from phage Lambda, causing self-lysis. Lysis released Perfringolysin O (*pfo*) and Phospholipase C (*plc*) from *Clostridium perfringens*, in turn causes rupture of the endosome. To demonstrate system functionality, fluorescent proteins were delivered to recipient mammalian cells. To enable *E. coli* to effectively kill tumor cells, we sought to identify an appropriate protein payload to deliver this system. We selected an endonuclease (BamHI), a ribonuclease (Barnase [111]), and a pore-forming hemolysin (HlyE [112]) for investigation. Expression of BamHI would be expected to irreparably damage the tumor cell DNA, thereby leading to cell death. The bacterial strain expressing BamHI carries the corresponding DNA methylase to prevent digestion of its own genome. Exposure to Barnase should lead to cleavage of cellular mRNA and subsequent cell death. To express Barnase in *E. coli*, it has been fused to a PelB export tag [113] and is expressed in a strain also expressing its inhibitor, Barstar, in the cytoplasm. Bacteria overexpressing HlyE have been previously reported to be toxic to mammalian cells [114]. Strains carrying the payload delivery device and each of these payloads were incubated with cancer cells *in vitro*, and their cytotoxicity assessed (Figure 4.3). While no killing was observed for BamHI or HlyE, substantial cellular detachment
was observed for cells exposed to Barnase expressing bacteria. We consider this to be evidence of bacterial mediated tumor cell killing.

Figure 4.4 – Bacterial treatment of tumors in mice. (a) Mice carrying tumors of implanted B16F10 cells were injected intravenously with either control bacterial (MC1061) or cytotoxic bacteria (JTK176 + BBa_J72211-BBa_J72198). Error bars represent the standard deviation of three replicates. (b) Mice carrying tumors of HCT116 cells were injected either intravenously (i.v.) or intratumorally (i.t.) with control bacteria (MC1061) or cytotoxic bacteria (JTK176 + BBa_J72211-BBa_J72198). Error bars represent the standard deviation of at least four replicates.

To follow up on the in vitro results, we sought to observe an effect on tumor growth in a mouse model. Mice bearing established subcutaneous B16F10 tumors were injected with bacteria carrying Barnase as a payload. However, no significant effect on tumor growth was observed (Figure 4.4a), and tumor sizes exhibited considerable variability. In an attempt to reduce the variability, a slower growing cell line, HCT116, was also tested, and in some cases bacteria were injected directly into the tumors to ensure colonization (Figure 4.4b). Although there may be a modest reduction in variability, the average tumor growth rates don not significantly differ, suggesting that further engineering is required to make E. coli an effective cancer therapeutic.

4.4 – Discussion

Our strategy for engineering a lab strain of E. coli to treat solid tumors entails making several modifications, including addition of protective capsular polysaccharides and incorporation of a direct cell-killing mechanism. The results presented here demonstrate substantial progress towards achieving both of those goals.

We have successfully modified a lab strain of E. coli to have either an O16 antigen or an O7 antigen, and demonstrated that they confer resistance to membrane mediated killing. For both O-antigens and the K1 capsule, omission of one of the required biosynthetic genes (wbbL,
manB, or neuS for O16, O7, and K1, respectively) proved to be an effective mechanism for controlling capsule expression. Unlike the O-antigens, however, we were ultimately unable to confer a protective K1 phenotype to a lab strain of \textit{E. coli}. Considerable effort was required to effect a genomic transfer of the biosynthetic genes from a hybrid strain, EV36, to a lab strain, MC1061. Even though the resulting strain exhibited sensitivity to a K1 specific phage, it was not resistant to other bacteriophage, unlike a natural K1 isolate where the O-antigen synthesis had been disabled. Similarly, attempts to clone the K1 capsule onto a plasmid failed, with poor complementation of K1 biosynthesis even in a natural isolate missing only a single K1 biosynthetic gene, and no apparent K1 biosynthesis in a lab strain. Although the mode of failure remains unclear, our difficulties working with K1 are not unique. A plasmid previously reported to confer a K1 phenotype (pRS23 [115]) failed to do so in our hands, and was described as “unstable” by the provider of the material. The polysialic acid biosynthesis cluster from \textit{Neisseria meningitidis} serogroup B is known to undergo frameshift mutations due to slipped-strand mispairing during replication [116], and it is possible that some similar mechanism leads to frequent loss of K1 cluster function in \textit{E. coli}. We suspect that recoding the K1 biosynthesis cluster may remove such obscure regulation and enable more controlled manipulation.

Although manipulation of capsule expression in a lab strain remains technically challenging, it stay may prove to be an effective approach for enabling therapeutic applications in humans. Even in a mouse model, where the complement is unable to directly kill \textit{E. coli} [117], the presence of a functional O-antigen and K-capsule was essential for maintaining substantially higher survival rates. We presume this extended lifetime would translate into more effective tumor localization, although this remains to be tested by exploring the dose of bacteria required to achieve robust localization behavior in encapsulated versus non-encapsulated strains.

Because even unprotected bacteria can localize to tumors in a mouse model, we were still able to investigate whether engineered \textit{E. coli} can affect tumor growth. We first assessed the ability of bacteria bearing both a toxic payload and a payload delivery device to kill cultured tumor cells \textit{in vitro}. Expression of a ribonuclease, Barnase, conferred a killing phenotype to \textit{E. coli}, although expression of an endonuclease, BamHI, and a hemolysin, HlyE, did not. These observations might be explained by more robust ability of mammalian cells to repair DNA damage[118], and an insufficient dose of HlyE to affect cell viability, although further investigation would be required to confirm those limitations. Because Barnase carries a periplasmic export signal sequence, it is possible that indirect tumor cell killing may have resulted from uptake of extracellular Barnase \textit{in vitro}. Confirmation of the mechanism of Barnase-mediated killing should be explored by omitting different components of the payload delivery device. Further, Barnase-mediated killing, as well as other killing mechanisms, should be explored with a range of bacterial concentrations, since a lower multiplicity of infection may help differentiate between more and less effective cytotoxic strategies. It may be the case that the relatively high concentrations of bacteria used to do the \textit{in vitro} studies here are not representative of the bacterial exposure of typical cells in a tumor.

After establishing that Barnase expressing bacteria are capable of cell killing \textit{in vitro}, we proceeded to test for an effect on tumor growth in a mouse model \textit{in vivo}. However, we observed no reduction in tumor growth rates after administering engineered \textit{E. coli}. There are a number of possible explanations for this failure, including poor distribution of \textit{E. coli} within the tumor, insufficient bacteria interacting with tumor cells to cause killing, and differences in bacterial behavior in a real tumor microenvironment. With \textit{Salmonella} \textit{spp.}, it was found that modification of the chemotactic machinery improved distribution throughout tumor tissue [119];
similar work may be required to improve \textit{E. coli} function. As described above, a high MOI was used to assess bacterial killing capability \textit{in vitro}, while only a few bacteria might interact with each tumor cell \textit{in vivo}. The payload delivery device was developed and tested on cancer cells \textit{in vitro}; it is conceivable that changes in the microenvironment of the tumor, such as reduced oxygen, could affect the behavior of the device \textit{in vivo}. Preliminary evidence suggests that none of the tumor cells receive the payload, implying that improving the robustness of the payload delivery device to conditions in tumor tissue should be explored. Such an investigation may be facilitated by the use of a 3D cell culture model, which more closely represents conditions in actual tissue [120-122].

Clearly, engineering \textit{E. coli} to act as a cancer therapeutic is an ambitious and challenging goal. Although we have made considerable progress toward this goal, several important obstacles remain. As both our technical ability to manipulate DNA at the scale of genomes [123, 124] and the fidelity of \textit{in vitro} assays improve, these obstacles should become easier to overcome.

\textbf{4.5 –Materials and Methods}

\textbf{4.5.1 –Plasmids, Strains, Phage, and Growth Media}

Plasmids used in this study are described in Chapter 6, and were constructed using BglBrick and BioBrick standard assembly [43, 125]. \textit{E. coli} strains MC1061 [45], MG1655 [107], EV36 [101], BOS12 [102], ATCC 23503 (referred to as O7:K1 natural isolate) or derivatives thereof were used for experiments. Derivatives were generated by the procedure of Datsenko and Wanner [62], by random mutagenesis [126], and by P1 transduction [127]. Full descriptions of plasmids and strain modifications are available in Chapter 6, and DNA sequences have been deposited in the Registry of Standard Biological Parts (http://partsregistry.org) [44]. Bacteriophages T7 and T4 were the kind gift of Richard Calendar. Bacteriohage K1E [104] was the kind gift of Dean Scholl. Bacterial cells were grown in Luria-Bertani medium (LB), or LB + 0.2% v/w arabinose for strains carrying pSB1A2-BBa_J72199, or LB + 40 mM MgSO$_4$ for strains carrying BBa_J72113-BBa_J72196 or BBa_J72218-BBa_J72197.

\textbf{4.5.2 – Serum Resistance}

Stationary phase cultures of strains to be tested were diluted to approximately 1 x 10$^5$ cfu/mL in LB. 5 µl of the diluted bacterial culture was added to both 100 µl of sterile filtered horse serum (Thermo Scientific # SH30074 03) and 100 µl of LB. After 1 hour incubation at 37°C, 50 ul of each mixture was plated for cfu determination. Bacteria were scored as resistant if the count of colonies from serum treated bacteria was at least 50% of the LB only treated bacteria, and scored as sensitive if the count of serum treated bacteria was less than 5% of the LB only treated bacteria.

\textbf{4.5.3 – Phage Sensitivity and Resistance}

Stationary phase cultures of strains to be tested were diluted 40-fold into LB and incubated for 1 hour at 37°C. Phage lysate was added, and bacteria incubated for an additional 1-2 hours at 37°C. Cultures were then inspected visually for clarification.
4.5.4 – Cell Culture Cytotoxicity

A monolayer of HeLa, B16F10 (ATCC: CRL-6475), or HCT116 (ATCC: CCL-247) cells was prepared 24 hours prior to experimentation in a 24-well culture plate in growth medium (DMEM or RPMI supplemented with 10% fetal bovine serum) with penicillin and streptomycin antibiotics. The medium was replaced with fresh medium without antibiotics, and either 3 µl of stationary phase bacterial culture or 1 µM (final concentration) staurosporine was added to each well, representing a multiplicity of infection of approximately 120. After 3 hours of incubation at 37 °C, cells were washed twice into growth medium with 100 µg/mL gentamicin, and the plate incubated for a further 24 hours. The cells were then washed once with DMEM before examination by microscopy. Detachment of most or all cells was considered to be evidence of cellular killing.

4.5.5 – Preparation of Bacteria for Animal Experiments

Bacteria were harvested at mid-logarithmic phase, mixed 1:1 with 50% glycerol, aliquoted into 1.5 mL microcentrifuge tubes, and frozen at -80°C. One aliquot was removed, the bacteria washed with PBS, and serial dilutions plated on LB-agar plates to estimate bacterial concentration. Prior to injection, a second bacterial aliquot was removed, washed with PBS, and resuspended at a final concentration of 1 * 10^8 cfu/ml. Serial dilutions of the final solution were plated on LB-agar plates to confirm the concentration of bacteria used for injection.

4.5.6 – Bacteria Clearance In Vivo

To assess bacterial survival, four- to six-week old female mice BALB/c mice (Jackson Laboratories) were injected into the lateral tail vein with 100 µl of bacterial suspension. After 120 minutes, 5 µl of blood was collected from the tail tip, immediately diluted into 100 µl of LB, and serial dilutions of the mixture plated on LB agar to determine cfu.

4.5.7 – Tumor Models and Bacterial Infections

Six- to seven-week-old male athymic nude mice (Jackson Laboratories Nu/J strain 002019) were injected subcutaneously on the right flank with 1.5 * 10^6 B16F10 murine melanoma cells (ATCC: CRL-6475) re-suspended in 100 µl PBS or 1.5 * 10^6 HCT116 human carcinoma cells (ATCC: CCL-247) re-suspended in 50 µl PBS and mixed 1:1 with growth factor reduced matrigel (BD Biosciences) for a total volume of 100 µl. Tumor-bearing nude mice were injected with bacteria after tumors reached a size of ~100-150 mm^3, typically around 1 week post cell implantation for B16F10 and 2 weeks post cell implantation for HCT116. Injections were either a single injection of 100 µl into the lateral tail vein or 2 injections of 50 µl directly into the tumor mass. Tumor volumes (calculated as L * W * W / 2) were monitored at least 3 times per week until they exceeded 1500 mm^3, at which point the study was terminated. To assess bacterial colonization of the tumor, mice were sacrificed, the tumors were excised and homogenized, and serial dilutions were plated on LB agar plates with or without appropriate single antibiotics. All animal experiments were carried out in accordance with protocols approved by the Animal Care and Use Committee (ACUC) of UC Berkeley.
Chapter 5 – Successes and Failures in Modular Genetic Engineering

Authors:

Joshua T Kittleson  
Department of Bioengineering, University of California, Berkeley CA, United States

Gabriel C. Wu  
Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA

J. Christopher Anderson (corresponding author: jcanderson@berkeley.edu)  
Department of Bioengineering, University of California, Berkeley CA, United States  
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley CA, United States


5.1 – Abstract

Synthetic biology relies on engineering concepts such as abstraction, standardization, and decoupling to develop systems that address environmental, clinical, and industrial needs. Recent advances in applying modular design to system development have enabled creation of increasingly complex systems. However, several challenges to module and system development remain, including syntactic errors, semantic errors, parameter mismatches, contextual sensitivity, noise and evolution, and load and stress. To combat these challenges, researchers should develop a framework for describing and reasoning about biological information, design systems with modularity in mind, and investigate how to predictively describe the diverse sources and consequences of metabolic load and stress.

5.2 – Introduction

Synthetic biologists engineer systems both to understand design principles relevant to natural systems (reviewed in [128]) and to address important environmental, clinical, and industrial needs (reviewed in [3, 4, 129]). Drawing on important concepts including abstraction, standardization, and decoupling [11], investigators have developed both novel systems that operate with little direct interaction with the host and synthetic systems intentionally embedded into native networks [130]. Here, we review how modular design, an important engineering strategy, has enabled development of more complex systems, and describe some of the failure modes that still challenge biological engineers, as well as offer strategies to address such design challenges.
5.3 – Modular Design

It was recognized several decades ago [131] that proteins are composed of modular domains, independently folding polypeptides that can be linked together to form a functional enzyme. More recently, biologists recognized that interacting collections of biomolecules that carry out discrete functions can also be seen as modules [132], and identified biological modules based on correlated protein activities [133-135]. These descriptive methods simplify analysis of native networks, but fail to provide a framework for predicting the behavior of engineered genetic networks. Although a formal theoretical framework for quantifying the robustness of a module to external module connections has been developed [136], it offers no insight into how to describe or characterize the biological operations performed by a module, or its in different cellular contexts. Absent theoretical guidance, biological engineers have focused on building a repertoire of modules with human defined functions [2] and characterizing modules in a way that allows their reuse [137].

Modular design envisions complex systems as compositions of functional modules that have been independently fabricated and characterized [138]. In mature engineering disciplines, module behavior is well-characterized, robust in expected use scenarios, and independent of other system elements. System design is reduced to specifying the necessary modules and connectivity needed to generate a desired higher order function. Currently, the characterization, reliability, and independence of biological modules are suspect [139]. The quantitative interrogation of biological systems is technically challenging and exacerbated by the fact that biological systems exist within the context of a cell. Some isolated biochemical entities such as green fluorescent protein (GFP) exhibit robust behavior across multiple contexts, but designing multicomponent modules that are sufficiently independent of cellular interaction remains a challenge.

Nevertheless, several groups have developed increasingly complex systems through the explicit use of modular design. In early work, Kobayashi and colleagues [140] framed system design as development of three modules: an input/sensing module, a regulatory/computing module, and an output/response module. Employing that framework, they engineered a cell-density dependent protein expression platform. They developed a sensing module using quorum sensing proteins from \textit{V. fischeri}, a computing module using a transcriptional regulator that behaves like a genetic toggle, and a response module that outputs GFP. Tabor and coworkers followed a similar strategy to develop a light-responsive genetic edge detection program [141]. They employed a previously developed dark sensor [142] as an input module, which was transcriptionally coupled to a regulatory module implementing edge detection logic. The transcriptional output of the edge detecting module served as an input to the response module, which expressed beta-galactosidase to enable visualization. The work illustrated how iterative testing of a system built up from constitutive components can enable successful development of a complex behavior.

As an alternative to describing expression levels as a signal that passes between genetic modules, metabolic engineers use metabolites as signals that pass between independently tested and fabricated biosynthetic modules. Martin and colleagues [143] engineered \textit{E. coli} to produce amorphadiene, a precursor to the antimalarial drug artemisinin, by first dividing the pathway into “top” and “bottom” operons. They constructed the bottom operon and confirmed production of amorphadiene from mevalonate, a stable intermediate that \textit{E. coli} can obtain exogenously, before providing the top operon and confirming function of the entire pathway. Similarly, Ajikumar and
coworkers [144] greatly improved biosynthesis of taxadiene, the first committed step towards synthesizing the anticancer drug Taxol, by separating the biosynthetic pathway into upstream and downstream modules. They joined the two modules by co-varying their expression and searching for a global optimum. Although the above examples demonstrate the potential for modular engineering, such systems are not generalizable, and the most elaborate synthetic systems developed to date are dwarfed by even a small natural developmental program [145]. To explain this gap, we consider some of the challenges relevant to module design, connection, and robustness.

Figure 5.1 – Failure modes of genetically engineered systems. (a) Syntactic errors, including truncations, mutations, and omissions, result in missing or incorrect components. (b) Semantic errors result from unintended or overlooked molecular interactions that adversely affect system or cellular function. (c) Mismatched parameters, such as imbalanced expression levels, can cause qualitative failures (e.g. dampened oscillations), fail to activate a desired phenotype, or trigger responses outside of specified conditions. (d) Changing the context of a system, as by changing the strain, copy number, or genomic location, can affect system behavior. (e) Populations of cells exhibit a variety of behaviors due to stochastic noise, and mutations may abrogate function entirely. (f) The additive metabolic load from one or more modules can manifest as either stress or toxicity.
5.4 –Failure Modes

5.4.1 – Syntactic Errors: Ensuring desired features are present

Function of biological modules and systems requires that all components are present in the DNA. However, user and database error can result in incorrect or missing components, called syntactic errors. While a powerful tool, automated gene annotation software can mis-annotate more than 10% of gene start sites [146, 147]. Naïve incorporation of an automatically-annotated open reading frame (ORF) into an engineered genetic system can either truncate the protein product, potentially abrogating function, or mislead designers about where to exert translational control.

5.4.2 – Semantic Errors: Getting the molecular interactions right

Semantic errors consist of unintended molecular interactions that adversely affect system function. For example, combining a module that employs a promoter with the widespread tet operator (tetO) sequence [148] as a constitutive source of transcription will likely fail if combined with a module that utilizes TetR (e.g. [149]). Similarly, introduction of a lox-based circuit (e.g. [150]) into a P1 lysogen will likely fail due to native Cre expression [151]. More difficult to anticipate than these well-characterized interactions are functions reported in the literature, but not widely cited. For example, CpG methyltransferases are toxic to wild type E. coli K12. Reports of cloning CpG methyltransferases explicitly cite the need to work in mcrA and mcrB strains [152-154], since those genes restrict methylated cytosines [155]. However, the mrr gene also restricts CpG methylated DNA [156] and while a mrr strain was used in two of the CpG methylase studies [152, 153], no comment about the need for this genotype appears in the text. The large and rapidly growing body of scientific literature makes it challenging and impractical to manually consider all possible interactions a priori and perform effective searches for information about those interactions.

Similarly difficult to anticipate are interactions that require a quantitative rather than qualitative description. In particular, cellular functions that are saturable or titratable will behave differently depending on the extent of their usage, and prediction of the switching point is challenging. For example protein degradation of SsrA-tagged proteins in prokaryotes normally follows Michaelis-Menten kinetics [157], but shift to a first order decay rate when the degradation machinery saturates. This phenomenon actually stabilized an oscillator system developed in E. coli [158]. The twin arginine translocation pathway can also be saturated by heterologous expression of tagged proteins, leading to accumulation of tagged proteins in the cytoplasm [159]. This may lead to the failure of a synthetic system relying on efficient export. Inversely, increased occupancy can affect system behavior. For example, introduction of lac operator binding sites on a multi-copy plasmid effectively reduces the concentration of available LacI regulator, enabling expression from an otherwise repressed genomic gene [160].

Unknown interactions are the most challenging for system designers. Substantial, but still incomplete, progress has been made mapping the metabolic, regulatory, and signaling networks of various organisms ([161-165]). Integration of heterologous elements exacerbates the problem. While engineering E. coli to overexpress the non-native metabolite mevalonate, toxicity from the accumulation of a pathway intermediate was observed [166]. Subsequent analysis demonstrated that the intermediate inhibited fatty acid biosynthesis, and that toxicity could be alleviated by
exogenous addition of palmitic acid [167] or scaffolding of pathway enzymes [168]. The scope of heterologous toxicity is hinted at by a recent survey of Sanger sequencing data, which indicated that over 10,000 sequenced genes from a variety of organisms have toxic interactions with *E. coli* [169].

One approach to minimizing unwanted interactions is segregation of system components away from other cellular entities. Localization to the inside of a protein shell offers a mechanism for accomplishing this segregation [170]. For example, the propanediol compartment from *C. freundii* can be expressed in *E. coli*, and desired proteins targeted to its interior [171]. Progress in the development of synthetic membrane bound organelles may ultimately offer a powerful method for compartmentalizing synthetic systems, preventing undesirable interactions [172].

### 5.4.3 – Mismatched Parameters: Tuning behaviors

Engineered systems will only exhibit the desired behavior if connected modules have matched input/output characteristics. In a process analogous to signal conditioning in electrical engineering, the output range of transcription or biomolecule production from one module must be in the range that causes a desired phenotype or triggers appropriate behavior in a downstream module. In early pioneering work, Yokobayashi and coworkers [13] illustrated this concept by connecting a genetic IMPLIES gate to an inverter gate. When the initial circuit failed to work properly, modification of the expression level or activity of the repressor carrying a signal between the two gates resulted in a functional design. Another group developed bacteria that release a toxic protein only when inside the anoxic environment of a tumor. Because the basal output level of the oxygen-sensitive promoter caused undesired synthesis of the toxic protein even in non-inducing conditions, the promoter was mutated to reduce its output level [82]. Similarly, Anderson and coworkers [173] engineered expression of invasin, which enables bacterial uptake into mammalian cells, to trigger only in the presence of a low oxygen or high bacterial concentration. To reduce undesired invasion in the absence of an appropriate signal, the RBS of invasin was mutated to reduce its basal expression. The paradigm of matching module activity is readily extended to engineering metabolic systems, where the importance of balancing pathway activity has been increasingly recognized [12]. In a recent ambitious engineering effort, simultaneous randomization of 20 ribosome binding sites in *E. coli* achieved a 390% increase in lycopene production [39]. Application of expression tuning is widespread in biosynthetic pathways and genetic networks. The recognition that appropriate coupling of modules and components is critical for establishing both qualitative and quantitative function has led to the development of several tools that allow relative expression levels of the same protein to be selected rationally [56-59].

### 5.4.4 – Contextual Sensitivity: Influence of factors beyond primary sequence

Changing the genetic context of a system also changes both the expression level and gene dosage. In mammalian systems, positional effects of different integration sites are known to alter expression levels [174]. When Kramer and colleagues constructed a hysteretic switch in mammalian cells, they integrated two sets of genes into separate loci [175]. The positional expression variability caused imbalanced production of system components in many screened clones, and prevented the desired switching behavior. Use of insulator elements can protect transgenes against both outside enhancers and gene silencing [176], as demonstrated by the improvement of transgene expression reliability in mice following inclusion of insulator
elements into the expression cassette [177]. In bacteria and yeast, transcriptional interference can alter expression [178], as can expression from upstream promoters. To enable more precise measurement of promoter activities in E. coli, promoter-probe vectors were created using terminators to insulate against transcription from both 5’ and 3’ elements [179]. Importantly, altering gene dosage by placing a system on a multi-copy plasmid both increases the level of gene expression and changes the number of cis DNA elements, such as operators or initiation sites. Theoretical work has demonstrated that behavior of a genetic circuit at single copy differs from the behavior at higher copy [180, 181]. Native systems employ feedback to mitigate the effects of gene dosage [42], inspiring development of an engineered system in yeast that uses feedback to ensure consistent protein expression levels regardless of the system copy number [182].

Environmental context changes, such as altered pH or temperature, can directly affect the activities of both biomolecules and small molecules. For example, acyl-homoserine lactone degradation rates vary with both pH and temperature [183], making systems employing quorum sensing potentially sensitive to corresponding environmental changes [184]. Activity changes in response to environmental shifts have also been deliberately incorporated into engineered systems [72-74].

5.4.5 – Noise and Evolution: System behavior in a population

Genetic systems exhibit behavioral diversity across a population, a consequence of both stochastic processes and the mutability of underlying genetic programs. Improvement in the ability to monitor the behavior of single cells has provided considerable insight into how stochastic processes generate noise in isogenic populations [185-187]. Fluctuations in the availability of shared resources, such as polymerase and ribosomes, has a global influence on expression levels, while the biochemical processes responsible for expression of individual genes causes further variation [188]. In prokaryotes, the balance between transcription and translation strength strongly influences expression noise, with low transcription and high translation generating the most variability [189]. In eukaryotes, bursts of transcription [190], potentially attributable to chromatin remodeling, also contribute to expression noise.

Variability in gene expression affects the behavior of several published synthetic systems. An engineered prokaryotic oscillator exhibited cell-to-cell variability in period and amplitude attributable to stochastic effects, requiring single cell observations to confirm oscillatory behavior [191]. In a eukaryotic regulatory switch constructed through positive feedback, stochastic variations in expression levels led to spontaneous state switching [192]. By contrast, natural systems both mitigate [185] and exploit [193, 194] noise and stochastic processes. A study of the E. coli chemotaxis pathway found that it has the smallest topology that offers sufficient noise suppression for detection of chemical signals [195], while B. subtilis maintains a competency decision network architecture that causes significant population heterogeneity as a bet hedging strategy [196]. As engineers, we can adopt strategies to similarly mitigate or exploit noise in the design of genetic systems. Coupling a slow, short-range interaction via a quorum sensing system to a fast, long-range interaction via a redox sensing system achieved the synchronized oscillation of millions of E. coli cells [197]. Elucidation of the principles behind engineering noisy promoters in yeast [198, 199] provides a tool for deliberate introduction of noise for probabilistic decision making.
Spontaneous genetic mutations arise in biological systems, introducing a confounding factor into genetic system design faced by few other engineering disciplines [11]. Although misbehavior of a single cell in large population is of little consequence when the desired system behavior relies on aggregate properties and requires little cellular growth, it can become problematic for both large-scale production and applications beyond the bioreactor [5]. Any burden imposed by an introduced genetic system (see discussion of load and stress below) could potentially be removed by a disabling mutation. Given enough growth cycles, the non-functional mutant will dominate the population [5]. While direct reduction of mutation rates due to DNA replication is challenging, elimination of mobile DNA elements can stabilize introduced genetic material. Systemic removal of insertion sequence elements from E. coli reduced mutation of toxic plasmids, enabling recovery of DNA with desired sequences [200]. While undesirable mutations usually remain problematic, engineers have taken advantage of spontaneous mutation and applied selective pressure to evolve desired behaviors. Computational analysis of a whole-genome metabolic reconstruction in E. coli identified a set of target genes that, upon removal, coupled bacterial growth to production of lactate. Subsequent adaptive evolution of the resulting strains by continuous growth substantially improved lactate titer [201].

5.4.6 – Chassis Failure: The role of load and stress

Indirect effects of an introduced system can alter the cellular state. Strong overexpression of an otherwise non-toxic protein causes competition for ribosomes and leads to cell death [202]. Moderate overexpression, can still effect the cellular environment, potentially resulting in growth inhibition, toxic accumulation of acetate, shutdown of the oxidative branch of the pentose phosphate pathway, reduced expression of housekeeping proteins, or stress responses [203]. Although phenomenological models exist to describe protein load’s effect on growth [204], consequences of cellular stress are not well understood. In yeast, for example, it has been predicted that decreased free ribosome concentration (as from expression of competing exogenous mRNAs) can lead to an ultrasensitive drop in translation of a subset of host mRNA [205]. Overexpression of membrane proteins also indirectly affects the host, likely saturating protein translocation machinery, consequently reducing respiratory chain complex formation and causing respiratory stress [206]. In cases where the biochemical action of a produced protein consumes energy [207] or subtly interacts with core host functions, the effect on the cellular environment is even less clear.

5.5 – The Path Forward

Natural systems often demonstrate remarkable modularity and robustness, with complex genetic programs [208, 209] and structures [210-212] readily transferrable between species. Focused effort in a few key areas can facilitate rapid progress towards that goal. First, knowledge about biological entities and systems, both natural and engineered, should be encoded in a computer-understandable format using a controlled vocabulary. Recognizing the need for codification in systems biology, several ontologies have been developed that enable description of system components, simulation and analysis methodology, and numerical results [213]. However, the synthetic biology community has yet to select or develop an ontology that meets the needs of biological engineers. Formal descriptions of current knowledge, and intelligent
inference algorithms that use it, are needed to scale designs beyond what is achievable through human intuition alone.

Because device engineering requires a significant time and resource investment, reuse of engineered components, modules, and systems is essential for rapid progress. Ideally, parts should be selected from large orthogonal sets, instead of reusing workhorses like TetR, to minimize collisions between devices [214]. For example, a set of RNA regulators was recently described in which rational changes to the RNA generates orthogonal variants [215]. Additionally, recent reductions to the cost of large-batch gene synthesis [216] will enable data mining of the metagenome for functionally equivalent but orthogonal sets of genetic components. Modules should also avoid dependence on host functions, just as broad host range plasmids and some phages encode most needed functions [217-220]. Devices not intended for bioproduction should minimize metabolic load on the host, avoiding unnecessarily draining components such as very strong promoters [221]. When possible, components should be placed at single copy on the genome flanked by insulating elements, thereby both reducing load and providing a more consistent gene dosage and expression level. Development of a framework for predictively assessing the effects of stress and load will be critical to avoiding system failure resulting from the additive effects of several modules. The synergistic effects of improvements to engineering methodology coupled with the advancement of core technologies such as DNA synthesis [10, 216, 222] and single cell manipulation and measurement [223-226] will enable development of increasingly complex and useful biological systems.

5.6 – Acknowledgements

This work was supported by the National Science Foundation Synthetic Biology Engineering Research Center (SynBERC). J.T.K. received support from a Siebel Scholar award. G.C.W. is supported by a National Science Foundation Graduate Research Fellowship.
# Chapter 6 – Plasmids and Strains

## Table 6.1 – Basic Parts

<table>
<thead>
<tr>
<th>Registry #</th>
<th>Short Description</th>
<th>Long Description/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBa_J27001</td>
<td>FRT</td>
<td>FLP recombinase recognition site, <em>Saccharomyces cerevisiae</em> [227]</td>
</tr>
<tr>
<td>BBa_J72005</td>
<td>P_{TET}</td>
<td>BglBrick standard version of BBa_R0040,TetR repressible promoter (see <a href="http://partsregistry.org/Part:BBa_R0040">http://partsregistry.org/Part:BBa_R0040</a>)</td>
</tr>
<tr>
<td>BBa_J72046</td>
<td>rbs.GFP.double terminator</td>
<td>BglBrick standard version of (BBa_E0040 joined with BBa_B0015)</td>
</tr>
<tr>
<td>BBa_J72048</td>
<td>rbs.sfGFP</td>
<td>Superfolder GFP mutant [228]. BglBrick standard version of BBa_I746916 (see <a href="http://partsregistry.org/Part:BBa_I746916">http://partsregistry.org/Part:BBa_I746916</a>)</td>
</tr>
<tr>
<td>BBa_J72049</td>
<td>double terminator</td>
<td>Tandem terminators; BglBrick standard version of BBa_B0015 (see <a href="http://partsregistry.org/Part:BBa_B0015">http://partsregistry.org/Part:BBa_B0015</a>)</td>
</tr>
<tr>
<td>BBa_J72050</td>
<td>P_{CON}-B5</td>
<td>BglBrick standard version of BBa_J72131</td>
</tr>
<tr>
<td>BBa_J72051</td>
<td>repB terminator</td>
<td>BglBrick standard version of BBa_J61048</td>
</tr>
<tr>
<td>BBa_J72052</td>
<td>rbs.brp!</td>
<td>A modified version of bacteriocin release protein from <em>E. coli</em>[229]</td>
</tr>
<tr>
<td>BBa_J72054</td>
<td>P_{pho}</td>
<td>Mg^{2+} responsive promoter from <em>E. coli</em>[230]</td>
</tr>
<tr>
<td>BBa_J72056</td>
<td>P_{nrP}</td>
<td>Mg^{2+} responsive promoter from <em>E. coli</em>[230]</td>
</tr>
<tr>
<td>BBa_J72059</td>
<td>terminator</td>
<td>BglBrick standard version of BBa_B1006, synthetic transcription terminator (see <a href="http://partsregistry.org/Part:BBa_B1006">http://partsregistry.org/Part:BBa_B1006</a>)</td>
</tr>
<tr>
<td>BBa_J72065</td>
<td>rbs_pelB&gt;</td>
<td>Leader sequence of pelB from <em>Erwinia carotovora</em> CE[231]</td>
</tr>
<tr>
<td>BBa_J72066</td>
<td>rbs_pelB_R&gt;</td>
<td>Leader sequence that exposes Arg at N-terminus after cleavage[232]. From <em>Erwinia carotovora</em> CE[231]</td>
</tr>
<tr>
<td>BBa_J72068</td>
<td>rbs_lamB_R&gt;</td>
<td>Leader sequence of lamB that expose Arg at N-terminus after cleavage[232]. From <em>E. coli</em>[233]</td>
</tr>
<tr>
<td>BBa_J72071</td>
<td>araC-P_{BAD}</td>
<td>Arabinose responsive regulator and corresponding promoter from <em>E. Coli</em> [234]</td>
</tr>
<tr>
<td>BBa_J72072</td>
<td>rhaRS-P_{BBA}</td>
<td>Rhamnose responsive regulators and corresponding promoter from <em>E. Coli</em> [235]</td>
</tr>
<tr>
<td>BBa_J72073</td>
<td>P_{CON}</td>
<td>Synthetic constitutive promoter</td>
</tr>
<tr>
<td>BBa_J72074</td>
<td>P_{CON-c5}</td>
<td>Synthetic constitutive promoter</td>
</tr>
<tr>
<td>BBa_J72075</td>
<td>P_{CON-g6}</td>
<td>Synthetic constitutive promoter</td>
</tr>
<tr>
<td>BBa_J72076</td>
<td>LP_{iu}</td>
<td>Late promoter, Bacteriophage P1 [236]</td>
</tr>
<tr>
<td>BBa_J72077</td>
<td>trupB terminator</td>
<td>Terminator from <em>E. coli</em> DH1</td>
</tr>
<tr>
<td>BBa_J72079</td>
<td>cin, no start, frameshift</td>
<td>Derived from Bacteriophage P1 [236]</td>
</tr>
<tr>
<td>BBa_J72080</td>
<td>repl., no start</td>
<td>Lytic replication origin, Bacteriophage P1 [236]</td>
</tr>
<tr>
<td>BBa_J72082</td>
<td>rbs.coi</td>
<td>Repressor inactivator, Bacteriophage P1 [236]</td>
</tr>
<tr>
<td>BBa_J72083</td>
<td>rbs.pacA</td>
<td>DNA packaging enzyme subunit and DNA packaging sites, Bacteriophage P1 [236]</td>
</tr>
<tr>
<td>BBa_J72084</td>
<td>rbs.mRFP1</td>
<td>Monomeric red fluorescent protein derivative [237]</td>
</tr>
<tr>
<td>BBa_J72085</td>
<td>rbs.repL</td>
<td>Initiates lytic replication, Bacteriophage P1 [236]</td>
</tr>
<tr>
<td>BBa_J72086</td>
<td>rbs.pir</td>
<td>Factor required for replication of gamma origin of <em>E. coli</em> plasmid R6K [28]</td>
</tr>
<tr>
<td>BBa_J72087</td>
<td>rbs.lacZ</td>
<td>Beta-galactosidase from <em>E. coli</em> MG1655</td>
</tr>
<tr>
<td>BBa_J72088</td>
<td>rbs.ColiE2 repA</td>
<td>Replication initiator of <em>E. coli/Shigella</em> plasmid ColE2-P9 [238], used to support replication of vector J72111 (referred to as pBJk2741 in [239])</td>
</tr>
<tr>
<td>BBa_J72089</td>
<td>rbs.cl</td>
<td>Master repressor, Bacteriophage P1 [236]</td>
</tr>
<tr>
<td>BBa_J72090</td>
<td>vioA BCDE</td>
<td>Synthetic (recoded) violacein synthesis operon from <em>Chromobacterium violaceum</em>. BglBrick standard version of BBa_K274002 (see <a href="http://partsregistry.org/Part:BBa_K274002">http://partsregistry.org/Part:BBa_K274002</a>)</td>
</tr>
<tr>
<td>BBa_J72091</td>
<td>luxCDABE</td>
<td>Luciferase and substrate synthesis operon from <em>Photorhabdus luminescens</em>(ATCC 29999) [240]</td>
</tr>
<tr>
<td>Registry</td>
<td>Description</td>
<td>Composition</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>BBa_J72092</td>
<td>kanamycin resistance cassette</td>
<td></td>
</tr>
<tr>
<td>BBa_J72128</td>
<td>&lt;plc!</td>
<td>Phospholipase C (plc) from Clostridium perfringens[241]</td>
</tr>
<tr>
<td>BBa_J72129</td>
<td>&lt;pfo!</td>
<td>Perfringolysin O (pfo) from Clostridium perfringens[241]</td>
</tr>
<tr>
<td>BBa_J72156</td>
<td>app 5’ UTR</td>
<td>MG1655</td>
</tr>
<tr>
<td>BBa_J72157</td>
<td>app 3’ UTR</td>
<td>MG1655</td>
</tr>
<tr>
<td>BBa_J72158</td>
<td>CoIE2 minimal origin</td>
<td>Minimal sequence thought to support replication from plasmid CoIE2-P9 [242]</td>
</tr>
<tr>
<td>BBa_J72159</td>
<td>rbs.repA</td>
<td>Replication initiation protein from plasmid CoIE2-P9 [242]</td>
</tr>
<tr>
<td>BBa_J72160</td>
<td>rbs2.repA</td>
<td>Replication initiation protein from plasmid CoIE2-P9 [242] with strong RBS</td>
</tr>
<tr>
<td>BBa_J72161</td>
<td>CoIE2 origin</td>
<td>Origin of replication from plasmid CoIE2-P9 [242]</td>
</tr>
<tr>
<td>BBa_J72162</td>
<td>rbs.pir</td>
<td>Replication initiation protein from E. coli R6K plasmid[27]</td>
</tr>
<tr>
<td>BBa_J72163</td>
<td>Pgyr</td>
<td>Promoter repressed by the presence of glucose [243]</td>
</tr>
<tr>
<td>BBa_J72164</td>
<td>dcr</td>
<td></td>
</tr>
<tr>
<td>BBa_J72165</td>
<td>rbs.dhfr</td>
<td>Dihydrofolate reductase (type II) from plasmid R-388[244]. Confers trimethoprim resistance.</td>
</tr>
<tr>
<td>BBa_J72166</td>
<td>rT2 terminator</td>
<td>A terminator from bacteriophage lambda with a &gt;90% termination efficiency.[245]</td>
</tr>
<tr>
<td>BBa_J72167</td>
<td>rbs.neuS (BBa)</td>
<td>Sialyltransferase responsible for chain elongation during synthesis of K1 capsule [246]</td>
</tr>
<tr>
<td>BBa_J72168</td>
<td>rbs.wbbL(BBa)</td>
<td>Rhamnosyltransferase required for O16 antigen synthesis [247]</td>
</tr>
<tr>
<td>BBa_J72169</td>
<td>O7 biosynthetic cluster</td>
<td>Pathway for synthesis of O7 antigen [248] from ATCC 23503</td>
</tr>
<tr>
<td>BBa_J72170</td>
<td>K1 biosynthetic cluster</td>
<td>Pathway for synthesis of K1 capsule [246] from ATCC 23503</td>
</tr>
<tr>
<td>BBa_J72171</td>
<td>rbs.neuS</td>
<td>Sialyltransferase responsible for chain elongation during synthesis of K1 capsule [246]</td>
</tr>
<tr>
<td>BBa_J72172</td>
<td>rbs.manB</td>
<td>Required for synthesis of GDP-mannose for O7 biosynthesis[248]</td>
</tr>
<tr>
<td>BBa_J72173</td>
<td>&lt;barnase!</td>
<td>Ribonuclease from Bacillus amylooliquefaciens[249]</td>
</tr>
<tr>
<td>BBa_J72175</td>
<td>rbs.barstar</td>
<td>Barnase inactivator from Bacillus amylooliquefaciens[249]</td>
</tr>
<tr>
<td>BBa_J72176</td>
<td>rbs.invasin.terminator</td>
<td>Invasin from Yersinia pseudotuberculosis [173]</td>
</tr>
<tr>
<td>BBa_J72178</td>
<td>rbs.malE&gt;</td>
<td>Malteose Binding Protein (encoded by malE) can be fused to other polypeptides and direct their export to the periplasm [250]</td>
</tr>
<tr>
<td>BBa_J72179</td>
<td>arac-PBAD (BBa)</td>
<td>Arabinose responsive regulator and corresponding promoter from E. Coli [234]</td>
</tr>
<tr>
<td>BBa_J72180</td>
<td>rbs.hlyE</td>
<td>Hemolysin protein from E. coli K12 [112]</td>
</tr>
<tr>
<td>BBa_J72181</td>
<td>rbs.NIS.R.BamHI</td>
<td>Restriction endonuclease BamHI [251] with a 4x repeat of the SV40 NLS [252]</td>
</tr>
<tr>
<td>BBa_K112305</td>
<td>rbs_ILysosome!</td>
<td>Lysozyme from enterobacteria phage λ[253]</td>
</tr>
<tr>
<td>BBa_K112311</td>
<td>rbs_lholin!</td>
<td>Holin from enterobacteria phage λ[253]</td>
</tr>
<tr>
<td>BBa_K112317</td>
<td>rbs_lantiholin!</td>
<td>Antiholin from enterobacteria phage λ[253]</td>
</tr>
<tr>
<td>BBa_R0040</td>
<td>PTEI (BBa)</td>
<td>TetR repressible promoter (see <a href="http://partsregistry.org/Part:BBa_R0040">http://partsregistry.org/Part:BBa_R0040</a>)</td>
</tr>
</tbody>
</table>

Table 6.2 – Composite Parts
<table>
<thead>
<tr>
<th>Index</th>
<th>Description</th>
<th>BBA References</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBa_J72098</td>
<td>Constitutively expressed cl</td>
<td>BBA_J72075.BBa_J72089</td>
</tr>
<tr>
<td>BBa_J72099</td>
<td>P_{BAD} driven, coi + repL + pacA phagemid for library fragment insertion</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72076.BBa_J72083.BBa_J72085</td>
</tr>
<tr>
<td>BBa_J72100</td>
<td>Constitutively expressed lacZ</td>
<td>BBA_J72074.BBa_J72087</td>
</tr>
<tr>
<td>BBa_J72101</td>
<td>P_{BAD} driven, coi + pacA phagemid</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72076.BBa_J72083.BBa_J72077</td>
</tr>
<tr>
<td>BBa_J72102</td>
<td>P_{BAD} driven, coi + cin phagemid</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72076.BBa_J72079</td>
</tr>
<tr>
<td>BBa_J72103</td>
<td>P_{BAD} driven, coi + repL + cin + pacA phagemid</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72079.BBa_J72080.BBa_J72076.BBa_J72083.BBa_J72077</td>
</tr>
<tr>
<td>BBa_J72104</td>
<td>P_{BAD} driven lax + vio operons, constitutive lacZ expression</td>
<td>BBA_J72072.BBa_J72091.BBa_J72090.BBa_J72049.BBa_J72074.BBa_J72087</td>
</tr>
<tr>
<td>BBa_J72105</td>
<td>Competitor phagemid; cin + repL + pacA</td>
<td>BBA_J72079.BBa_J72080.BBa_J72083</td>
</tr>
<tr>
<td>BBa_J72106</td>
<td>P_{BAD} driven, coi + RFP phagemid</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72005.BBa_J72084.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72107</td>
<td>Reference promoter driving expression of GFP</td>
<td>BBA_J72047.BBa_J72046</td>
</tr>
<tr>
<td>BBa_J72108</td>
<td>P_{BAD} driven GFP</td>
<td>BBA_J72071.BBa_J72046</td>
</tr>
<tr>
<td>BBa_J72115</td>
<td>P_{BAD} driven, coi + cin + repL phagemid</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72079.BBa_J72080</td>
</tr>
<tr>
<td>BBa_J72116</td>
<td>P_{BAD} driven, coi + cin + pacA phagemid</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72079.BBa_J72076.BBa_J72083.BBa_J72077</td>
</tr>
<tr>
<td>BBa_J72152</td>
<td>P_{BAD} driven, coi + repL + cin + pacA phagemid expressing GFP</td>
<td>BBA_J72071.BBa_J72082.BBa_J72049.BBa_J72079.BBa_J72080.BBa_J72076.BBa_J72083.BBa_J72077.BBa_J72005.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72155</td>
<td>GFP expression cassette</td>
<td>BBA_J72005.BBa_J72048.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72182</td>
<td>repA expression cassette</td>
<td>BBA_J72005.BBa_J72159</td>
</tr>
<tr>
<td>BBa_J72183</td>
<td>GFP expression cassette</td>
<td>BBA_J72163.BBa_J72048</td>
</tr>
<tr>
<td>BBa_J72184</td>
<td>repA expression cassette for genomic integration</td>
<td>BBA_J72073.BBa_J72160.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72185</td>
<td>GFP expression cassette</td>
<td>BBA_J72005.BBa_J72048.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72186</td>
<td>Trimethoprim resistant, pir expression cassette for genomic integration</td>
<td>BBA_J72156.BBa_J72177.BBa_J72162.BBa_J72049.BBa_J72001.BBa_J72073.BBa_J72165.BBa_J72166.BBa_J72001.BBa_J72157</td>
</tr>
<tr>
<td>BBa_J72187</td>
<td>dxs containing sequence</td>
<td>BBA_J72048.BBa_J72159.BBa_J72164</td>
</tr>
<tr>
<td>BBa_J72188</td>
<td>weak violacein expression cassette</td>
<td>BBA_J72074.BBa_J72090</td>
</tr>
<tr>
<td>BBa_J72189</td>
<td>strong violacein expression cassette</td>
<td>BBA_J72005.BBa_J72090</td>
</tr>
<tr>
<td>BBa_J72190</td>
<td>newS expression cassette (BBa)</td>
<td>BBA_r0040.BBa_J72167</td>
</tr>
<tr>
<td>BBa_J72191</td>
<td>wblL expression cassette (BBa)</td>
<td>BBA_r0040.BBa_J72168</td>
</tr>
<tr>
<td>BBa_J72192</td>
<td>manB and newS expression cassette</td>
<td>BBA_J72005.BBa_J72172.BBa_J72171</td>
</tr>
<tr>
<td>BBa_J72193</td>
<td>manB expression cassette</td>
<td>BBA_J72005.BBa_J72172</td>
</tr>
<tr>
<td>BBa_J72194</td>
<td>newS expression cassette</td>
<td>BBA_J72005.BBa_J72171</td>
</tr>
<tr>
<td>BBa_J72195</td>
<td>Barnase/Barstar expression cassette</td>
<td>BBA_J72073.BBa_J72065.BBa_J72173.BBa_J72077.BBa_J72174.BBa_J72175.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72196</td>
<td>Payload Delivery Device</td>
<td>BBA_J72056.BBa_J72052.BBa_K112305.BBa_K112311.BBa_J72049.BBa_J72050.BBa_K112317.BBa_J72051.BBa_J72056.BBa_J72066.BBa_J72129.BBa_J72068.BBa_J72128.BBa_J72059</td>
</tr>
<tr>
<td>BBa_J72197</td>
<td>Payload Delivery Device</td>
<td>BBA_J72054.BBa_J72052.BBa_K112305.BBa_K112311.BBa_J72049.BBa_J72050.BBa_K112317.BBa_J72051.BBa_J72056.BBa_J72066.BBa_J72129.BBa_J72068.BBa_J72128.BBa_J72059</td>
</tr>
<tr>
<td>BBa_J72198</td>
<td>Payload Delivery Device with Barnase Payload</td>
<td>BBA_J72054.BBa_J72052.BBa_K112305.BBa_K112311.BBa_J72049.BBa_J72050.BBa_K112317.BBa_J72051.BBa_J72056.BBa_J72066.BBa_J72129.BBa_J72068.BBa_J72128.BBa_J72059.BBa_J72074.BBa_J72178.BBa_J72065.BBa_J72173.BBa_J72077.BBa_J72174.BBa_J72175.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72199</td>
<td>wblL newS expression cassette (BBa)</td>
<td>BBA_J72179.BBa_J72168.BBa_J72167</td>
</tr>
<tr>
<td>BBa_J72200</td>
<td>hlyE expression cassette</td>
<td>BBA_J72074.BBa_J72180.BBa_J72049</td>
</tr>
<tr>
<td>BBa_J72201</td>
<td>BamHI endonuclease expression cassette</td>
<td>BBA_J72005.BBa_J72181</td>
</tr>
</tbody>
</table>
Table 6.3 – Plasmid Backbones

<table>
<thead>
<tr>
<th>Registry #</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBa_J72109</td>
<td>ampicillin resistant pUC vector</td>
</tr>
<tr>
<td>BBa_J72110</td>
<td>ampicillin + chloramphenicol resistant p15A vector</td>
</tr>
<tr>
<td>BBa_J72111</td>
<td>spectinomycin resistant split ColE2 vector</td>
</tr>
<tr>
<td>BBa_J72112</td>
<td>spectinomycin resistant split R6K vector</td>
</tr>
<tr>
<td>BBa_J72113</td>
<td>chloramphenicol resistant p15A vector</td>
</tr>
<tr>
<td>BBa_J72114</td>
<td>complete, arabinose inducible phagemid, chloramphenicol resistant, p15a vector</td>
</tr>
<tr>
<td>BBa_J72153</td>
<td>ampicillin + kanamycin resistant p15A vector</td>
</tr>
<tr>
<td>BBa_J72154</td>
<td>kanamycin + chloramphenicol resistant p15A vector</td>
</tr>
<tr>
<td>BBa_J72202</td>
<td>ampicillin resistant p15A vector</td>
</tr>
<tr>
<td>BBa_J72203</td>
<td>kanamycin resistant split R6K vector</td>
</tr>
<tr>
<td>BBa_J72204</td>
<td>ampicillin resistant split R6K vector with O16 homology and kanamycin resistance flanked by FRT sites</td>
</tr>
<tr>
<td>BBa_J72205</td>
<td>ampicillin resistant split ColE2 vector</td>
</tr>
<tr>
<td>BBa_J72206</td>
<td>ampicillin resistant split R6K vector</td>
</tr>
<tr>
<td>BBa_J72207</td>
<td>ampicillin resistant pUC vector</td>
</tr>
<tr>
<td>BBa_J72208</td>
<td>Kanamycin resistant BAC with R6K origin (BBa format)</td>
</tr>
<tr>
<td>BBa_J72209</td>
<td>kanamycin + ampicillin resistant p15A vector</td>
</tr>
<tr>
<td>BBa_J72210</td>
<td>Kanamycin resistant BAC with R6K origin, glucose driven promoter S’ of cloning site</td>
</tr>
<tr>
<td>BBa_J72211</td>
<td>ampicillin resistant p15A vector with <em>dapD</em></td>
</tr>
<tr>
<td>BBa_J72212</td>
<td>spectinomycin resistant pUC vector</td>
</tr>
<tr>
<td>BBa_J72218</td>
<td>kanamycin resistant p15A vector</td>
</tr>
<tr>
<td>pSB1A2</td>
<td>ampicillin resistant pUC vector (BBa format)</td>
</tr>
</tbody>
</table>

Table 6.4 – Strains

<table>
<thead>
<tr>
<th>Name</th>
<th>Description/Genotype</th>
<th>Composition of insertions</th>
<th>Registry #</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCC 23503</td>
<td>O7:K1 isolate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS12</td>
<td>O16:K92 isolate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EV36</td>
<td>Hybrid K1-K12 strain derived from CGSC288 and RS1085 (which was derived from D699) [101]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JH5-18</td>
<td>MC1061 P21::P&lt;sub&gt;nr&lt;/sub&gt;-inv [90J]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JTK011</td>
<td>ATCC 23503 Δ&lt;i&gt;neuS&lt;/i&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JTK029</td>
<td>MC1061 upp::pir+FRT-kanR-FRT Cassette encoding constitutive expression of pir and kanamycin resistance inserted between purM and &lt;i&gt;uraA&lt;/i&gt;</td>
<td>purM - BBa_J72005.BBa_J72086.BBa_J72001.BBa_J72092.BBa_J72021 - &lt;i&gt;uraA&lt;/i&gt;</td>
<td>BBa_J72136</td>
</tr>
<tr>
<td>JTK030</td>
<td>MC1061 upp::pir+FRT Cassette encoding constitutive expression of pir inserted between purM and &lt;i&gt;uraA&lt;/i&gt;</td>
<td>purM - BBa_J72005.BBa_J72086.BBa_J72001 - &lt;i&gt;uraA&lt;/i&gt;</td>
<td>BBa_J72137</td>
</tr>
<tr>
<td>JTK035</td>
<td>ATCC 23503 Δ&lt;i&gt;manB&lt;/i&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JTK037</td>
<td>ATCC 23503 Δ&lt;i&gt;manB ΔneuS kanR&lt;/i&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JTK041</td>
<td>ATCC 23503 Δ&lt;i&gt;manB ΔneuS&lt;/i&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JTK050</td>
<td>ATCC 23503 Δ&lt;i&gt;manB ΔneuS trans&lt;/i&gt;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
JTK160  MC1061 O16::repA+FRT-kanR-FRT  galF - BBa_J72160.BBa_J72049.BBa_J27001.BBa_J72092.BBa_J27001-wbbL  BBa_J72213

JTK160C  MC1061 O16::repA+FRT  Cassette encoding constitutive expression of repA (from CoLE2) and kanamycin resistance inserted between galF and wbbL  galF - BBa_J72073.BBa_J72088.BBa_J72049.BBa_J72001.BBa_J72092.BBa_J27001 - wbbL  BBa_J72138

JTK164  MC1061 upp::pir+FRT-trimethoprim-R-FRT  purM - BBa_J72177.rbslib-BBa_J72162.BBa_J72049.BBa_J27001.BBa_J72073.BBa_J72165.BBa_J72166.BBa_J27001-uraA  BBa_J72214

JTK165EI  MC1061 upp::rbsE.pir+FRT-trimethoprim-R-FRT O16::rbs1repA+FRT  purM - BBa_J72177.rbsE-BBa_J72162.BBa_J72049.BBa_J27001.BBa_J72073.BBa_J72165.BBa_J72166.BBa_J72001-uraA  galF - BBa_J72073.rbs1-BBa_J72160.BBa_J72049.BBa_J27001-wbbL  BBa_J72214

JTK174  MC1061 ΔdapD O16::P_glpT-rbs1.inv  galF- BBa_J72163.(rbs1)BBa_J72176.BBa_J27001-wbbL (composition is approximate - see sequence in registry)  BBa_J72216

JTK176  MC1061 ΔdapD O16::P_glpT-rbs2.inv  galF- BBa_J72163.(rbs2)BBa_J72176.BBa_J27001-wbbL (composition is approximate - see sequence in registry)  BBa_J72217

MC1061  F araD139 MarA9-leu7697 Δ(codB-lacI)3 Δ(lac)X74 galK16 galE15: e14 mcrA0 relA1 rpsL150(strR) spT1 mcrB1 hsdR2

MC828E  MC1061 O16(wbbL) K1(ΔneuS)

MC828Q  MC1061 O16(wbbL) K1(ΔneuS) ΔmsbB Δfin ΔtonB

MG1655  F λ ilvG rfb-50 sph-1

49
Chapter 7 – References


