

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**RELIABLE GENE EXPRESSION AND ASSEMBLY FOR
SYNTHETIC BIOLOGICAL DEVICES IN *E. COLI*
THROUGH CUSTOMIZED PROMOTER INSULATOR
ELEMENTS AND AUTOMATED DNA ASSEMBLY.**

by

SWATI BANERJEE

MS in Cell & Molecular Biology, University of Rhode Island, 2009
BSc in Biotechnology, Bangalore University, 2005

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2016

© Copyright by
SWATI BANERJEE
2016

Approved by

First Reader

Douglas M Densmore, PhD
Associate Professor of Electrical & Computer Engineering

Second Reader

Jake Beal, PhD
Scientist, BBN Technologies

Third Reader

Wilson Wong, PhD
Assistant Professor of Biomedical Engineering

Acknowledgments

First and foremost I would like to thank my PhD advisor and major professor, Dr. Douglas M. Densmore for giving me the opportunity to pursue my doctoral dissertation in his lab. I joined his lab in unusual circumstances and I will forever remain grateful for the chance he gave me when others did not. In addition to providing me with a lab and all the supplies and equipment necessary for my research, he allowed me complete freedom in the choice of a research topic and trusted me to steer my project as my interests and curiosity directed me. When I needed help that was beyond his area of expertise, he connected me with the best scientists in those fields. He encouraged me to step outside the confines of the lab and my specific research and engage with the broader scientific community as a conference participant, conference organizer and in scientific outreach efforts. Not only was he tolerant of my explorations in science education at levels other than university, he actively helped me find and pursue such opportunities. I cannot imagine a more dynamic doctoral research experience possible than the one I received as part of Doug's group.

I would also like to thank Dr. Jacob Beal, who has filled the roles of collaborator, mentor and committee member. As a collaborator he ran analyses on the data I collected from my experiments. As a mentor, he helped me design my experiments and provided invaluable feedback on my manuscript and thesis, and patiently answered my many, many questions. I would also like to thank the rest of my committee members - Professors Ahmad Khalil, Wilson Wong and Trevor Siggers for their feedback to my work over the years.

None of my work would have been possible without the constant and continuing support of my loving husband, Keith. I would not have made it this far without his

unwavering faith in my ability even when I did not share that faith. I would like to also thank my parents who informed me at the age of 7 that I would be pursuing a PhD, even though I did not quite know what that meant then. They instilled in me my love of knowledge and learning. Although they neither know nor care, I would like to thank my two cats, Kaiserin Shivani and Rupert Baxter for keeping me sane and being the “audience” to all early runs of talks and poster presentations that I have given throughout my doctoral work.

My acknowledgments would be incomplete if I did not thank Dr. Traci Haddock; she has been a supervisor and a friend. She is the first person I call when an experiment goes either right or wrong, or when I have two hours between steps in an experiment and there is an oyster happy-hour at a restaurant downtown. In the same vein, I would like to thank Dr. Swapnil Bhatia, Janoo Fernandes and Vanessa Yanez for being the most amazing, entertaining and quirky set of friends I could have asked for.

This work was supported by the National Science Foundation Award number 1253856. This work was also supported by the Office of Naval Research (ONR) under award number N000141110725.

**RELIABLE GENE EXPRESSION AND ASSEMBLY FOR
SYNTHETIC BIOLOGICAL DEVICES IN *E. COLI*
THROUGH CUSTOMIZED PROMOTER INSULATOR
ELEMENTS AND AUTOMATED DNA ASSEMBLY.**

(Order No.)

SWATI BANERJEE

Boston University, Graduate School of Arts and Sciences, 2016

Major Professor: Douglas M. Densmore, PhD, Associate Professor

ABSTRACT

Building reliable genetic devices in synthetic biology is still a major challenge despite the various advances that have been made in the field since its inception. In principle, genetic devices with matching input and output expression levels can be assembled from well-characterized genetic parts. In practice, *a priori* genetic circuit design continues to be difficult in synthetic biology due to the lack of foundational work in this area. Currently, a successful genetic device is typically created by manually building and testing many combinatorial variants of the target device and then picking the best one. While this process is slow and error-prone, as synthetic genetic devices grow in complexity, this approach also becomes unmanageable and impractical.

Fluctuations in genetic context have been identified as a major cause of rational genetic circuit design failures. Promoter elements often behave unpredictably as they are moved from the context in which they were originally characterized. Thus, the ordered location of parts in a synthetic device impacts expected performance.

Synthetic spacer DNA sequences have been reported to successfully buffer promoters from their neighboring DNA sequence but design rules for these sequences are lacking.

I address this problem with a novel method based on a randomized insulator library. I have developed a high-throughput, flow cytometry-based screen that randomly samples from a library of 4^{36} potential insulators created in a single cloning step. This method provides precise control over genetic circuit expression. I further show that insulating the promoters in a genetic NOT-gate improves circuit performance and nearly eliminates the effect of the order in which the promoters are organized in the device. This foundational work will help improve the design of reliable genetic devices in *E. coli*.

Finally, automated DNA assembly using liquid-handling robots can help increase the speed at which combinatorial synthetic device variants are assembled. However, these systems require significant investment in optimizing the handling parameters for handling very small volumes of the various liquids in DNA assembly protocols. I have optimized and validated these liquid-handling parameters on the Tecan EVO liquid handling robotic platform. These materials have been made available to the larger community.

Contents

1	Introduction	1
2	Background Information	8
2.1	Synthetic biology: definition and goals	8
2.2	Modern synthetic biology vs. traditional genetic engineering	9
2.3	Top-down vs. bottom-up synthetic biology	11
2.3.1	Engineering principles in synthetic biology and an analogy to computer engineering	12
2.4	Genetic Context in Synthetic Biology	13
2.4.1	Compositional context as a source of circuit performance failures	14
3	Materials and Methods	18
3.1	Materials	18
3.1.1	Media	18
3.1.2	Bacterial strains	18
3.1.3	Antibiotics	19
3.1.4	Kits	19
3.1.5	Enzymes	19
3.1.6	DNA parts and primers	20
3.2	Methods	21
3.2.1	DNA purification	21

3.2.2	Modular Cloning (MoClo) DNA assembly	22
3.2.3	Randomized spacer generation and screening	24
3.2.4	Colony PCR	26
3.2.5	Transformation and selection	26
3.2.6	DNA sequencing	27
3.3	In vivo assay methods	28
3.3.1	Growth of bacterial cells	28
3.3.2	Fluorescence measurements by flow cytometry	28
3.3.3	Molecules of Equivalent Fluorescein (MEFL) conversion	28
4	DNA Spacers: Preliminary Work	41
4.1	Introduction	41
4.2	Problem Statement: Synthetic genetic device performance is dependent on the order of intermediate transcription units	45
4.3	Promoter Architecture for uninsulated promoters J23100, pBAD and pTet	46
4.3.1	Uninsulated J23100 promoter architecture	46
4.3.2	Uninsulated pBAD promoter architecture	47
4.3.3	Uninsulated pTet promoter architecture	48
4.4	Characterization of promoter expression in isolation	49
4.4.1	Inverter component TU promoters are affected by 4 bp MoClo fusion sites immediately upstream of promoter	49
4.4.2	Neutral DNA sequences do not reliably insulate promoters from upstream local DNA sequence context	52
4.4.3	Rationally designed “neutral” 36nt DNA spacer sequences affect promoter expression	53

4.4.4	Expanding the rationally designed J23100_A 36nt DNA spacer library did not yield DNA spacers providing consistent insulation	54
4.4.5	Rationally designed DNA spacers provide unreliable insulation	55
4.5	Conclusions	56
5	DNA Spacers: High-throughput screening of insulated promoter libraries	71
5.1	Introduction	71
5.2	Results	72
5.2.1	Distribution of 36nt-insulated J23100 expression cassettes . . .	72
5.2.2	Analysis of DNA spacer sequences from J23100_X screens . . .	76
5.2.3	Distribution of PCR-generated 36nt-insulated pBAD library . .	77
5.2.4	Distribution of PCR-generated 36nt-insulated pTet library . .	78
5.3	Discussion	80
5.3.1	Designing DNA spacers of alternate lengths	80
5.3.2	Resource Cost of Insulating a Promoter	81
5.3.3	Transferring the randomized spacer screening methodology to other organisms	83
5.3.4	Generating improved context-independent variants of commonly used BioBricks promoters	84
6	DNA Spacers: Component order independence and precise expression control through the use of empirically selected DNA spacers	90
6.1	Introduction	90
6.2	Results	91

6.2.1	Empirically insulating constitutive promoter J23100 in 24 inverter set improves overall device performance and reduces DNA sequence-based order dependence	91
6.2.2	Insulating inducible promoter pBAD in 24 inverter set improves overall device performance and reduces DNA sequence-based order dependence.	92
6.2.3	Co-insulating promoters J23100 and pBAD in the 24 inverter set eliminates upstream promoter DNA sequence-based order dependence	92
6.2.4	Insulating all promoters with custom-screened DNA spacers improves overall device performance and increases the fraction of devices performing as expected	98
6.3	Discussion	100
7	DNA Assembly Automation: Development and validation of liquid classes for automated DNA assembly on a Tecan liquid-handling robotic platform	107
7.1	Introduction	107
7.1.1	Existing automation in synthetic biology	108
7.1.2	Current state of DNA assembly automation	109
7.2	Optimizable Liquid Class Parameters	111
7.2.1	Liquid Class Parameters	111
7.2.2	Global Liquid Class Settings	118
7.3	Liquid Class Optimization	119
7.4	Experimental validation of optimized liquid classes	123
7.4.1	Automated DNA Assembly	124

7.4.2	Modular Cloning (MoClo) Reaction Parameters and reaction protocol:	125
7.4.3	Automated DNA assembly protocol and setup	126
8	Conclusions and Future Work	129
8.1	Replacing minimal constitutive promoters with consistent constitutive promoters	129
8.1.1	Promoter Insulation Workflow	131
8.1.2	Applying DNA spacer screening in other organisms	135
8.1.3	Mining DNA spacers for an improved DNA spacer development process	136
9	Appendix A: 24 Inverter Permutations Raw Data	144
10	Appendix B: PERL Script for DNA Spacer Primer Design for MoClo-format promoters	157
	Bibliography	161
	Curriculum Vitae	168

List of Tables

3.1	Bacterial growth media used in this study	18
3.2	Antibiotics used in this study.	19
3.3	Molecular kits used in this study.	20
3.4	Enzymes used in this study.	21
3.5	Basic Modular Cloning (MoClo) DNA parts and vectors used to construct higher level genetic circuits.	30
3.6	Promoter parts insulated with 12 bp spacers.	31
3.7	Promoter parts insulated with 24 bp spacers.	32
3.8	Promoter parts insulated with 36nt spacers.	33
3.8	Promoter parts insulated with 36nt spacers.	34
3.9	List of 36nt insulated promoters used in this study. The following promoters were created using DNA spacers selected using my screening methodology as described in Chapter 3, Section 3.3 and are available through Addgene [1]. Insulated promoter parts will also be submitted to the Registry of Biological Parts [35].	35
3.10	List of primers used for sequencing.	35
3.11	List of primers used to generate insulated promoter libraries.	36
4.1	Composition characteristics for rationally designed DNA spacer sequences.	55

5.1	Approximate reagent cost to insulate a promoter	82
5.2	Approximate time cost to insulate a promoter	82
7.1	Aspirate parameters for enzyme liquid class by volume-based sub-classes	120
7.2	Stepwise optimization for aspirate protocol for enzyme liquid class. . .	120
7.3	Dispense parameters for enzyme liquid class by volume-based sub-classes	121
7.4	Aspirate Parameters for DNA liquid class by volume-based sub-class .	122
7.5	Dispense Parameters for DNA liquid class by volume-based sub-class	123
7.6	Modular Cloning (MoClo) reaction setup	126
7.7	Experimental validation of optimized liquid classes	127
9.1	GFP and RFP expression of uninsulated 24 inverter permutations upon induction with 1mM L-Arabinose.	145
9.2	GFP and RFP expression of uninsulated 24 inverter permutations in the absence of induction with 1mM L-Arabinose.	146
9.3	GFP and RFP expression of 24 inverter permutations with all ratio- nally insulated promoters upon induction with 1mM L-Arabinose. . .	147
9.4	GFP and RFP expression of 24 inverter permutations with all ratio- nally insulated promoters in the absence of induction with L-arabinose.	148
9.5	GFP and RFP expression of promoter J2300-only insulated 24 inverter permutations upon induction with 1mM L-Arabinose.	149
9.6	GFP and RFP expression of promoter J23100-only insulated 24 in- verter permutations in the absence of induction with 1mM L-arabinose.	150
9.7	GFP and RFP expression of inducible promoter pBAD-only insulated 24 inverter permutations upon induction with 1mM L-Arabinose. . .	151

9.8	GFP and RFP expression of inducible promoter pBAD-only insulated 24 inverter permutations in the absence of induction with 1mM L-arabinose.	152
9.9	GFP and RFP expression of promoters J23100 and pBAD-insulated 24 inverter permutations on induction with 1mM L-arabinose.	153
9.10	GFP and RFP expression of promoter J23100 and promoter pBAD-insulated 24 inverter permutations in the absence of L-Arabinose induction	154
9.11	GFP and RFP expression upon induction with 1mM L-arabinose of the 24 inverter set with all promoters (J23100, pBAD and pTet) insulated.	155
9.12	GFP and RFP expression in the absence of induction with 1mM L-arabinose of the 24 inverter set with all promoters (J23100, pBAD and pTet) insulated.	156

List of Figures

2.1	Computer engineering-inspired hierarchy analogy for biological organization	12
2.2	Changing 4 base pairs at the 5 boundary of promoters alters promoter expression	17
3.1	Construction of a Level 1 MoClo part.	37

3.1	Construction of a Level 1 MoClo part. (Continued from previous page.) By allowing matching fusion sites between the 3' end of the first part and the 5' end of the second part, and so on, this ensures that the parts assemble in the desired assembly order of promoter-RBS/5'-UTR/CDS-terminator. The DV fusion sites match the fusion sites on the 5' end of the promoter and the 3' end of the terminator to allow the properly assembled transcriptional unit to be cloned into the vector in the correct orientation. The assembled transcriptional unit replaces the <i>lacZ</i> fragment and colonies containing kanamycin-resistant plasmids with assembled transcriptional unit grow on antibiotic media with IPTG and X-GAL as white colonies, thus enabling blue-white screening along with antibiotic selection. Figure reproduced with permission from Type IIS Assembly for Bacterial Transcriptional Units: A Standardized Assembly Method for Building Bacterial Transcriptional Units Using the Type IIS Restriction Enzymes BsaI and BbsI by Traci Haddock <i>et al</i> , 2015 from BioBricks Foundation. [30]	38
3.2	Construction of a Level 2 MoClo part.	39
3.3	Schematic for the design, construction and screening of randomized spacer library.	40
4.1	Bacterial transcriptional inverter (NOT-gate) architecture:	43
4.2	Expression profile of the 24 uninsulated inverter permutations of our test TetR inverter upon induction with L-arabinose.	58

4.2	Expression profile of the 24 uninsulated inverter permutations of our test TetR inverter upon induction with L-arabinose.	
	(Continued from previous page.) Assembly scars upstream of promoters cause anomalous expression and order dependence of modules in genetic devices assembled using Golden Gate-derived DNA assembly methods. Comparison of GFP and RFP fluorescence fold changes upon induction with L-arabinose of the 24 inverter permutation set in which no promoters are insulated reveals device-to-device variations and a clear pattern of high GFP fluorescence in inverters where pBAD-GFP_A was TU1 in the inverter configuration. All fluorescence readouts were converted to absolute units (MEFLs). Conversion of fluorescence readouts to MEFLs was done using the TASBE tools by our collaborators.	59
4.3	J23100 promoter architecture	59
4.4	pBAD promoter architecture	60
4.5	Model of AraC induction by L-arabinose at the araBAD promoter (PBAD)	61
4.6	pTet promoter architecture	62
4.7	Expression of J23100_X GFP cassettes in isolation	63
4.8	Expression of pBAD_X GFP expression cassettes in isolation	64
4.9	Expression of pTet_X RFP expression cassettes in isolation	65
4.10	RFP expression from promoter cassettes insulated with 12nt, 24nt and 36nt DNA spacer sequences	66
4.11	Evaluating the effect of DNA spacer sequences on promoter expression level.	67

4.12	Expanding the rationally designed 36nt DNA spacer library showed variability in insulated J23100_A expression	68
4.13	Expression profile of 24 inverter set with all promoters insulated with rationally selected DNA spacers upon induction with 1mM L-arabinose.	69
4.13	Expression profile of 24 inverter set with all promoters insulated with rationally selected DNA spacers upon induction with 1mM L-arabinose. (Continued from previous page.)	
	Rationally designed DNA spacers do not guarantee reliable insulation. Comparison of GFP and RFP fluorescence fold changes upon L-arabinose induction of 24 inverter permutations built with rationally designed insulated promoters removes previously observed high GFP fold change pattern from the uninsulated inverter set, but does not eliminate device to device expression fluctuations. All fluorescence readouts were converted to absolute units (MEFLs).	70
5.1	Distribution of insulated 36nt-J23100_A GFP library.	73
5.2	Distribution of insulated 36nt-J23100_E GFP library.	74
5.3	Distribution of insulated 36nt-J23100_H GFP library.	76
5.4	Distribution of insulated 36nt-J23100_K GFP library.	85
5.5	Distribution of DNA spacer nucleotide bases from insulated J23100_ libraries	86
5.6	Sequence overlap distribution for all sequenced DNA spacers.	87
5.7	Distribution of pBAD_F GFP library insulated with 36nt DNA spacers	88
5.8	Distribution of pTet_A RFP library insulated with 36nt DNA spacers	89

6.1	Expression profile of 24 inverter set with promoter J23100 insulated empirically upon induction with 1mM L-arabinose.	93
6.1	Expression profile of 24 inverter set with promoter J23100 insulated empirically upon induction with 1mM L-arabinose. GFP and RFP fold change upon L-arabinose induction of J23100-insulated 24 inverter set. DNA spacer sequence was obtained by screening the 36n-J23100_E library against J23100_A. Insulated promoters are denoted in the device key by small gray boxes placed between the purple box denoting J23100-AraC and its upstream MoClo fusion site. Insulation of J23100 resulted in a smaller fraction of both devices that fail or behave anomalously. Interestingly, out of the 8 devices that either failed or had an expression pattern violating the NOT-gate truth table, 6 devices contained pBAD_F with either GFP or TetR as the CDS. All fluorescence readouts have been converted to absolute units (MEFLs).	94
6.2	Expression profile of the 24 inverter set with promoter pBAD insulated empirically upon induction with 1mM L-arabinose.	95

6.2	Insulating promoter pBAD with screened DNA spacer reduces context dependence and improves overall inverter performance. GFP and RFP fold change upon L-arabinose induction of pBAD-insulated 24 inverter set. DNA spacer sequence was obtained by screening the 36n-pBAD_F library against pBAD_A. Insulated promoters are denoted in the device key by small gray boxes placed between the upstream MoClo fusion site and the blue and green boxes denoting pBAD-TetR and pBAD-GFP TUs, respectively. Insulation of J23100 resulted in a smaller fraction of both devices that fail or behave anomalously compared to the uninsulated 24 inverter set. All fluorescence readouts have been converted to absolute units (MEFLs).	96
6.3	Expression profile of the 24 inverter set with promoters J23100 and pBAD insulated empirically upon induction with 1mM L-arabinose. .	102
6.3	Expression profile of the 24 inverter set with promoter all promoters insulated empirically upon induction with 1mM L-arabinose.	104

6.3	Expression profile of the 24 inverter set with promoter all promoters insulated empirically upon induction with 1mM L-arabinose. (Continued from previous page.) Insulating the upstream DNA context-independent promoter pTet improves inverter performance. GFP and RFP fold change upon L-arabinose induction of J23100, pBADand pTet-insulated 24 inverter set. DNA spacer sequences were obtained from the 36n-J23100_E, 36n-pBAD_F and 36n-pTet_A screens. Insulated promoters are denoted in the device key by a small gray box placed between the purple, blue, green or red box denoting J23100-AraC, pBAD-TetR, pBAD-GFP or pTet-RFP, respectively, and its upstream MoClo fusion site. Selecting a DNA spacer from 36nt-pTet screen that provided a >10x repression on L-arabinose induction improves the RFP fold change in the 24 inverter set.	105
6.4	Observed incremental improvement in GFP and RFP expression of the 24 inverter set upon insulating with DNA spacers selected using screening methodology.	106
7.1	Screenshot of Tecan Dispense Liquid Class Panel	116
7.2	Screenshot of Tecan Aspirate Liquid Class Panel	124
7.3	Setup for Reagent, DNA parts and MoClo reaction plates in automated DNA assembly.	128
8.1	Decision process workflow for promoter insulation	138
8.2	High-level schematic for promoter insulation process	139
8.3	Simplified decision process workflow for promoter insulation	140

8.4	Promoter insulation protocol applied to constitutive minimal promoter J23100.	141
8.5	Promoter insulation protocol applied to constitutive minimal promoter J23100.	142
8.6	Promoter insulation protocol applied to repressible promoter J23100.	143

Chapter 1

Introduction

Synthetic biology holds the potential to alter the way humans interact with their environment, providing customized genetic solutions to fields ranging from bioremediation, fuel generation and fabric production all the way to personalized medical diagnostics and treatment. Through the application of principles from engineering disciplines, synthetic biology aims to transform biology into a systematic and predictable discipline that allows users to take genetic parts from diverse sources from off-the-shelf kits, mix and match them using the knowledge contained within part datasheets and then assemble them into novel, functioning genetic networks without having to go through laborious trial and error design and assembly cycles.

While genetic parts are modular in principle, meaning they can be swapped in and out of genetic circuits to alter circuit behavior, being composed of DNA that is at the same time the building material for genetic parts and the chemical code that controls part behavior, genetic parts are also sensitive to changes in DNA sequence composition at their boundaries. Swapping genetic parts in and out can alter the local chemistry at part junctions, or alter the distance between critical regulatory regions where proteins such as RNA polymerase or ribosomes bind. Although rare, there is the possibility that in fusing two genetic parts, a new part is created at the part:part junction that has a deleterious effect on circuit performance. These

complications pose challenges to rational genetic circuit design in spite of detailed component part characterization data being available. Successful genetic circuits are often arrived at by brute force: synthetic biologists are relying heavily on exhaustive assembly and testing of all combinatorial variants of a genetic circuit, from which the best performing one is selected [71]. This design strategy is slower and impractical for scaling up in size and complexity [58]. So, while synthetic biology has the potential to deliver solutions to a multitude of real-world problems, in order to deliver on its promises, synthetic biology must deliver cost-effective alternatives to current technologies in order to be commercially viable options.

The contribution of part-junction interference due to altered DNA sequence context at part boundaries has now been adequately documented and is widely accepted as a leading cause of failed *a priori* circuit design [13, 11]. While early synthetic biology focused heavily on making large numbers of genetic parts from diverse sources available to encode new behaviors and provide control over expression levels, the focus now is shifting towards making genetic parts truly modular. Novel strategies using the latest technologies have been developed and new part collections have been made available to remove the context dependence of Ribosome Binding Sites (RBSs) at both the RBS:promoter and RBS:gene boundaries in bacterial genetic circuits [50, 45, 59]. Although the effect of fluctuating DNA sequence at the 5' boundary of bacterial promoters is recognized, much less effort has been devoted towards removing these effects, possibly because the DNA sequence immediately upstream of a genetic circuit is not technically a part of the engineered circuit and contains no active genetic elements.

Genetic devices are typically assembled by combining component parts together in the order in which the genetic logic is executed. In higher-level devices composed of

intermediate expression modules, fusing modules creates new part:part junctions that did not exist in the original context in which component parts were characterized. In the past few years, synthetic biologists have been paying a closer attention to the effect of these part junction DNA sequence fluctuations and their effect on overall circuit performance.

The ability of certain DNA sequences approximately 40 to 100 bases upstream of the transcription start site (TSS) to increase the rate of transcription from a promoter has been known for years [21, 47]. More recently, the strong, and extremely variable effect of altered DNA sequences at the immediate 5' boundary of promoters has been documented [38].

We have also observed, however, that due to unexpected effects of DNA context at the 5' boundaries of promoters, an alternative configuration of a complex device can function where the version assembled conventionally fails entirely. Assembling and testing all possible permutations of a genetic device is a cumbersome endeavor, and is rarely done in practice. Moreover, due to part-junction interference, a promoter rationally selected for a device based on its available characterization data may behave differently when placed into a device with different DNA bases at its 5' boundary. Thus far, fusing mostly arbitrarily selected, synthetic insulator sequences to the 5' ends of promoters is the only available solution to part-junction interference due to genetic context fluctuations at the upstream boundaries of bacterial promoters. Their use is far from widespread, which may be because of a combination of the paucity of well-characterized insulated promoter parts, and the fact that insulator elements being themselves composed of DNA, could - and most likely would - themselves cause fluctuations in promoter performance that have not been characterized and therefore cannot be modeled. Scanning prior work [16] revealed that there were

not enough instances of insulated promoters in bacterial genetic circuits to develop a set of rules for designing reliable insulators. Only a single length of DNA spacer was tested, and the goal of insulation in that work was to buffer the promoter part from DNA sequences -40 to -60 base pairs upstream of the transcription start site. The work presented in this thesis studies DNA spacers as a tool to buffer promoters against DNA sequence changes at the 5' junctions of promoters.

Chapter 4 begins with a depiction of how part-junction interference at the 5' boundary of promoters leads to dependence on the order of component transcription units (TUs) on circuit performance. As my design, I chose a 4-TU inverter circuit regulated entirely in trans, and created all possible tandem permutations of the 4 component TUs to create 24 inverters that are equivalent in design. I then demonstrate my preliminary attempt to determine the optimal length and design characteristics of DNA spacer elements for insulating bacterial promoters from the 4 bp cloning scars present at the 5' promoter boundary as a result of DNA assembly using the Type IIS restriction endonuclease-based, Modular Cloning (MoClo). As the addition of a stretch of DNA sequence would serve primarily to add spacing between the promoter and the upstream MoClo assembly scar that we knew to be causing fluctuations in expression, I refer to these elements in the body of this work as DNA spacers rather than as DNA insulators to make the distinction that they lack specific insulating properties present in eukaryotic transcriptional machinery in which insulators naturally function. I investigate 3 lengths - 12nt, 24nt and 36nt - of DNA spacers upstream of 4 commonly used promoters, 3 of which are constitutive minimal while the fourth is a repressible promoter. The selected DNA spacer sequences were randomized sequences with no stable secondary structures or protein-binding consensus sequences. Additionally, their DNA base composition fell

within the accepted GC% of the intergenic regions between *E. coli* operons. Finding that 36nt DNA spacers appear to reliably insulate two of the four tested promoters (both of which are present in my inverter circuit mentioned above), while the 12nt and 24nt DNA spacers did not, I proceeded to expand my DNA spacer library so that I could ultimately create a library of insulated promoter parts that could be added to public part repositories for common use. However, the expression of the expanded insulated promoter library fluctuated from one DNA spacer to the next. Moreover, I found these DNA spacers to be ineffective in insulating the promoters in my abovementioned inverter circuit to remove upstream sequence-based component order dependence.

What was needed then, was a new methodology that would allow me to screen potential DNA spacer sequence candidates and select ones that actually served as insulators and/or spacers: sequences that either create enough space between the promoter and potentially problematic DNA sequences upstream, or, when inserted upstream of a promoter that is already within the final context of DNA sequence in which it must function, corrects whatever DNA sequence-based performance variations that have arisen due to reorganization of the promoter into a new surrounding DNA sequence. I then present the data obtained through the high-throughput screening of the insulated promoter libraries. Sticking with the 36nt length of DNA spacer that provided partial success in the preliminary work, to facilitate rapid insulated promoter generation, I developed an inverse PCR-based method that would create a library of potentially 4^{36} insulated promoters in a single cloning step using constitutive promoter J23100 that I had successfully insulated in my preliminary work in Chapter 4.

In Chapter 5, I employ a novel, high throughput, flow cytometry-based screening

method in which candidate insulated promoter cassette expression was compared against a reference device of known expression level or desired expression level. I also demonstrate the application of my screening method on an inducible promoter (pBAD) and a repressible promoter (pTet) promoter. The strength of my novel DNA spacer screening method is that it allows one to screen for a DNA spacer that provides precise expression control empirically in the desired DNA sequence context without the need for additional modeling. I also show that my DNA spacer screening method can be applied to inducible and repressible promoters.

In Chapter 6, I validate the utility of my method in the ultimate goals of eliminating variations in promoter expression due to part-junction interference at the 5' boundaries of promoters in my 24 inverter permutation set. I do this by showing that the incremental application of my method to context-affected promoters eliminates order dependence and improves overall performance. Finally, I apply my method to an unaffected promoter to further improve the overall device performance of my 24 inverter permutations.

When the goal is not to perfect the expression of a single device, but instead to produce large numbers of variations on a circuit, possibly for the creation of a library, it is more suitable to perform combinatorial DNA synthesis. Automated DNA assembly using liquid-handling robots is an efficient way to do this. However, outside of commercial ventures, DNA synthesis is mostly done manually. Although this is partly due to the relatively higher cost of liquid-handling robotics setup as well as higher cost of resources/reaction, automated DNA synthesis suffers from the lack of defined parameters for handling the very small volumes of liquids of various viscosities (especially enzymes). Automated single-reaction scale DNA assembly on liquid-handling platforms is inefficient when it works at all. In Chapter 6, I outline the

liquid-handling parameters for aspirating and dispensing $2\mu\text{L}$ volumes of enzymes, water. These parameters provide $>99\%$ efficiency of DNA assembly using the MoClo DNA assembly technique. These parameters were arrived at through extensive trial and error; all parameters were tested by calculating the number of true positive colonies in the automated construction of a constitutive GFP expression cassette (with manual assembly as a control at each step). Deviations from these parameters result in 0-1% cloning efficiency. I verify that the liquid handling parameters can be successfully applied to create 16 single MoClo transcription units from basic MoClo parts.

My contributions to the field of synthetic biology are as follows:

1. I have developed and verified a novel method for eliminating DNA sequence based context effects at bacterial promoters that will reduce the reliance on trial and error in the assembly of functioning genetic circuits.
2. I have applied that method to a constitutive promoter, an inducible promoter, and a repressible promoter.
3. Using my insulated promoters, I have eliminated component order dependence of transcription units within a higher level genetic circuit
4. I have deposited 3 new upstream genetic context-proof promoter elements to the standardized part repository, Addgene (and will also be adding them to the Registry of Standard Biological Parts).
5. I have developed a set of liquid-handling parameters for automated DNA assembly on a Tecan liquid-handling robotic platform that will enable automated DNA synthesis without the addition of additional expensive robotic attachments to pipettors.

Chapter 2

Background Information

2.1 Synthetic biology: definition and goals

Over the years, synthetic biology has represented a broad spectrum of endeavors involving the modification of naturally occurring biological systems. Synthetic biology has encompassed anything from the design of enzymes and *in vitro* systems, to the decoupling of native regulation of environmentally stimulated promoters to allow greater control over their regulation and all the way to creating entirely synthetic replicating life forms, each of which requires genetic manipulation. However, what sets modern synthetic biology apart from traditional genetic engineering is the application of engineering principles like abstraction, standardization, modularity and predictive modeling to the process of bioengineering.

Much of synthetic biology research today aims to transform biology into a systematic, predictable engineering discipline. Synthetic biology of today is often understood as the application of engineering principles to biology with the aim of encoding natural systems and their components with novel, programmable behaviors. Applying such principles, naturally occurring logical expressions from one or multiple organisms are reorganized and recombined to specify new, artificial ones. To achieve these ends, synthetic biologists often design new genetic parts, *in vitro* expression systems, biological and non-biological platforms for testing engineered ge-

netic circuits, predictive models for evaluating component choices for genetic circuit design, programming languages to convert logical expression specifications into valid biological circuitry, develop new DNA assembly techniques for rapid, combinatorial DNA assembly as well as computational tools for planning complicated assemblies and finally establish metrology guidelines for evaluating the performance of engineered genetic circuits. In doing so, synthetic biologists also seek to gain a deeper, more fine-grained understanding of the principles guiding the rewiring of the genetic components to make all the above processes more accurate and efficient.

2.2 Modern synthetic biology vs. traditional genetic engineering

For much of its history, biology and engineering occupied largely separate realms within the world of scientific inquiry with little, if any, overlap. Biology had to be accepted and understood while engineering created and controlled. The perplexing complexity of biological systems, both within a single system, and as present throughout the phylogenetic tree, defied the controlled organization of engineered devices. Both disciplines thrived on their own separate trajectories.

In the past 60 years, scientists have made enormous advances in biology, particularly in molecular biology, and have succeeded in breaking down the vast stretches of chemically monotonous DNA molecules into many smaller defined, modular units, specifying their boundaries, describing their characteristics, and mapping them to specific functions. In 1961, Jacob and Monod published their work on the lac operon describing how bacteria use transcription to regulate their behavior in response to an external signal (lactose) [39]. Descriptions of other transcriptional regulation networks soon followed. The discovery of the tools of genetic engineering restric-

tion endonucleases and DNA ligase enabled Herbert Boyer and Stanley Cohen in 1973 to create the first ever recombinant DNA molecule in the laboratory. Since then, molecular biologists have been engineering biology at the lab bench: In 1973 a mouse was genetically modified animal (a mouse) by inserting DNA from the Simian Virus 40 (SV40) into an early stage mouse embryo and showing that the viral DNA was present in every cell of the adult mouse. Since then, genetic engineering has flourished giving birth to the biotechnology industry where microbes are used as protein factories for the production of insulin, growth hormone, or clotting factors in the medical world, for modifying crops to make them more nutritious (golden rice), or more pest-resistant or drought resistant, and for the commercial production of enzymes (alpha-amylase, chymosin etc.).

However, most traditional genetic engineering efforts and achievements involve only the insertion of one or a few genes within existing natural genetic frameworks in order to achieve upregulation of their protein products [58]. Modularity in biology was under-utilized. Synthetic biology recognizes the modularity in biology at all layers of biological hierarchy, from basic genetic parts all the way to the level of biological pathways. Coupled with abstraction of function between layers, synthetic biology aims to reorganize the syntax of genetic machinery to encode entirely new cellular outcomes. Furthermore, synthetic biology attempts to change the *ad hoc* nature of bioengineering design and replace it instead with an off-the-shelf genetic parts toolkit and clear design principles that eliminate the guesswork and the mystique behind engineering biology.

2.3 Top-down vs. bottom-up synthetic biology

Synthetic biology can often be viewed in terms of top-down and bottom-up approaches to engineering biology. The top-down approach seeks to identify the minimum complement of genes required for survival, then strip down the living system of all redundancies. Synthetic biologists have redesigned the bacteriophage T7 genome by replacing almost 30% of its natural genetic elements with engineered DNA. In 2008, the first completely chemically assembled genome was reported [27], and was subsequently used as the only genetic material inside a fully synthetic bacterial cell [28]. The Craig Venter Institute has successfully identified the essential genes in *Mycoplasma genitalium* [29, 33]. Another group reduced the genome of *E. coli* K-12 by as much as 15% and found the resulting genome-reduced *E. coli* to possess higher electroporation efficiency and to accurately propagate recombinant DNA that was found to be unstable in other *E. coli* strains [57]. Forster et al are building a minimal cell that contains its own replication and essential biochemical pathways, but requires only small molecule nutrients for its survival [23]. More recently, in 2014, scientists at Johns Hopkins University have created a synthetic, truncated version of the *Saccharomyces cerevisiae* chromosome III and showed it to be fully functional in yeast [5].

The construction of artificial genetic circuits from scratch with basic elements can be considered the bottom up approach to synthetic biology. This approach uses our current understanding of genetic circuitry to remove parts from their native circuitry (decoupling) and re-purpose them, or to use synthetic part libraries. In the bottom-up approach, synthetic biologists design and construct new biological parts, which are used to build genetic circuits. These circuits link sensing an external stimulus with the execution of engineered logic (digital or analog) with a measurable

output. Making synthetic genetic circuits and devices *de novo* from basic parts enables synthetic biologists to better understand the rules that govern circuit design and composition. The bulk of synthetic biology research being done is bottom up synthetic biology.

The work presented in this thesis focuses on bottom-up synthetic biology.

2.3.1 Engineering principles in synthetic biology and an analogy to computer engineering

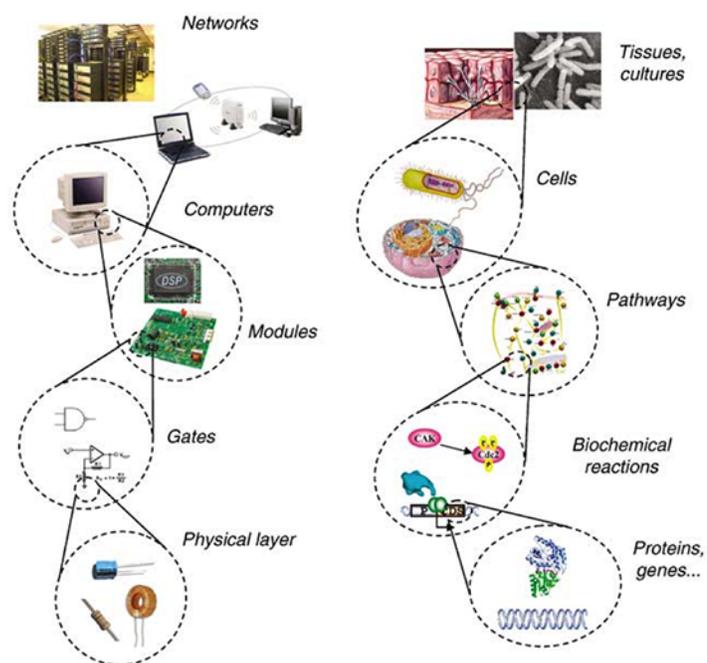


Figure 2.1: **Computer engineering-inspired hierarchy analogy for biological organization.** Andrianantoandro *et al.* drew parallels between the organization of electronic components in a hierarchy of increasing complexity to form systems and then networks and genetic parts and the complex regulatory networks and biological systems that are formed by their organization. Figure reprinted from Synthetic biology: new engineering rules for an emerging discipline, by Andrianantoandro *et al.*, Molecular Systems Biology, 2006. [4]

Using the concept of modularity and abstraction, Andrianantoandro *et al.*[4],

demonstrate synthetic biology as analogous to electrical and computer engineering in its hierarchical organization: proteins and genes form the bottom-most layer of the hierarchy analogous to the physical layer of components such as capacitors, resistors etc., in electrical engineering. Just as the physical layer of components can be combined to form logic gates, genetic regulatory elements and protein coding sequences can be organized to produce genetic circuit that carry out biochemical reactions. In electrical engineering, modules are created by combining and layering individual logic gates; similarly, the output of individual biochemical reaction-producing genetic circuits can be layered by having the output of one feed into another as its input and to create pathways. Just as a computer contains multiple modules, cells contains multiple pathways. Computers themselves can be connected to each other to form networks. Similarly, cells are organized to communicate with each other as part of tissues in multicellular organisms. Complexity in biology arises when modules at various layers interact with each other [2], but are handled through decoupling and abstraction.

Simplifying biological systems into modular abstracted layers allows synthetic biologists to investigate, through bottom-up construction, biology from its very first principles. The bulk of synthetic biology research done today is the design, construction and testing of bottom-up synthetic genetic circuits.

2.4 Genetic Context in Synthetic Biology

While borrowing analogies from other engineering disciplines has been beneficial to the development of the synthetic biology paradigm, engineering analogies do not transfer to biology perfectly. The majority of engineered genetic circuits fail to behave according to predictions in spite of rigorous modeling of performance based

on extensive characterization of component parts and modules [45, 13]. This often leads to a breakdown in rational iterative design in favor an exhaustive, resource-intensive approach of combinatorial synthesis of all variants of a target genetic device. Often a genetic circuit that is found to perform as expected in one host organism (chassis) fails to express at all when transferred to a different host organism.

Unlike electrical and computer systems, biological systems are complex, in constant flux and frequently not orthogonal to the component parts within the synthetic genetic circuit. Even the most well-studied host organism, *E. coli*, is not fully understood, and much of its biochemistry is not within the synthetic biologists' control. Furthermore, the synthetic circuit being tested can place strains on the host organisms' growth. The synthetic circuits themselves utilize host resources, and often the components of synthetic circuits are also part of the host genetic machinery. In both cases, it becomes difficult to determine the effect of the host on the performance of the genetic circuit. The physical composition of component parts into genetic circuits also alters genetic context at the part junctions, leading to unexpected and unpredictable part and device expression in the final context.

2.4.1 Compositional context as a source of circuit performance failures

At the circuit level, a fundamental mechanism of controlling circuit behavior is by swapping genetic regulatory elements such as promoters and RBSs to achieve fine control of expression. Altering the promoter:RBS pair or the RBS:gene pair alters the DNA sequence at part:part junctions, which has been found to sometimes and unexpectedly impact how a part functions, resulting in a type of compositional genetic context effect known as part-junction interference [13, 11]. Complex, multi-layered genetic devices can be created by connecting the output of one circuit to the input

of the next [49]. Each connection creates new part-part junctions that are potential sites for unexpected genetic context effects that alter circuit behavior.

Compositional context arising from part-junction interference is one of the leading causes of *a priori* circuit design. Part-junction interference undermines the fundamental principle of composability of independently characterized genetic parts and part modules into predictably behaving larger genetic devices, as well as restricting the utility of characterization information of genetic parts in part repositories.

2.4.1.1 Promoter context

Promoters in particular have been reported to be sensitive to changes in DNA sequence at both their 5' and 3' sequence boundaries. Altering the DNA sequence at the promoter:RBS junction can change the transcription start site [41, 67], which in turn can alter the length and sequence of the 5' untranslated region (UTR), affecting promoter melting and polymerase escape frequency [16]. Cleaving the 5' UTR to standardize RBS accessibility via ribozymes [45] or CRISPR processing [59], or promoter:RBS junctions using insulated RBS parts have all proven to be effective strategies for eliminating local context effects 3' of the promoter in bacterial systems [50].

Short promoter parts (<50 bp) have been reported to be susceptible to altered local context upstream of the promoter elements [3, 43, 64, 40, 32, 32, 46]. The α -domain of RNA polymerase (RNAP) can contact the DNA sequence approximately 100 bp upstream of the promoter [51] and alter its activity.

A recent study found that a 4 bp sequence immediately upstream of the -35 RNAP binding site of bacterial minimal constitutive promoters can alter promoter expression by up to 13x [38]. Synthetic insulator sequences have been shown to

relieve some local compositional context effects by standardizing DNA sequence at the 5' boundary of the promoter [16, 25]. One study created a small library of insulated minimal bacterial promoters using a single insulator sequence [16]. However, I show in this thesis that rationally designed insulators of similar base composition themselves alter promoter expression variably and cannot be reliably used without being extensively characterized (See Results section in Chapter 4). Furthermore, long-range and off-target effects of an insulator DNA sequence on other parts within a larger genetic device cannot be predicted.

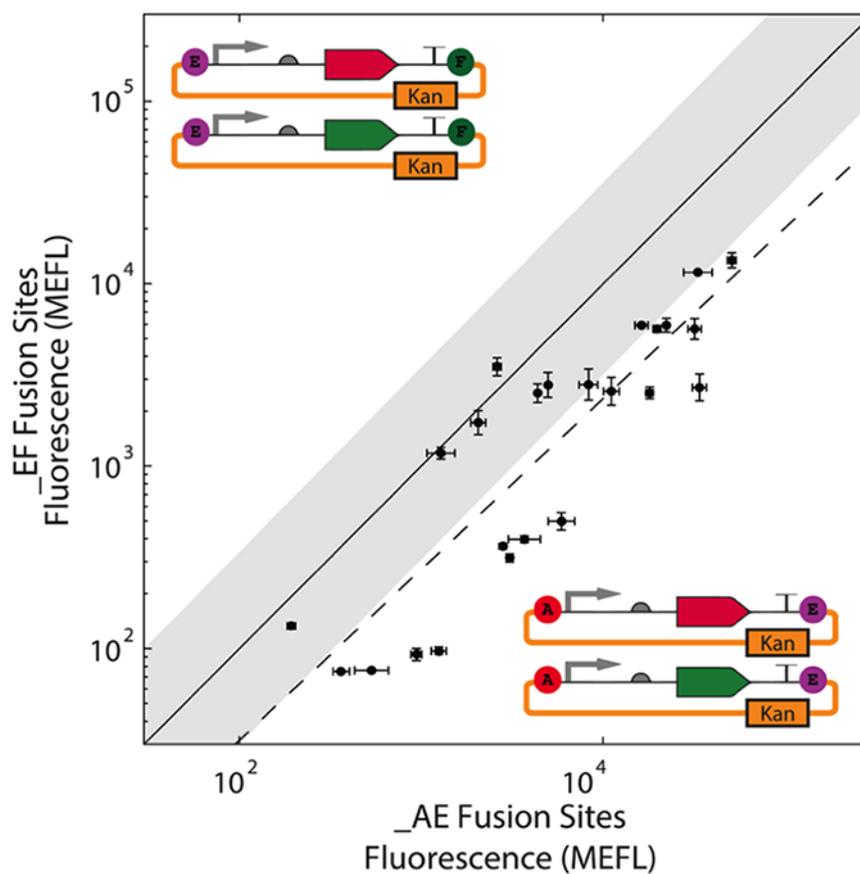


Figure 2.2: **Changing 4 base pairs at the 5' boundary of promoters alters promoter expression.** Changing the 4 base pair scar at the 5' boundary of a given transcription unit influences expression likely due to the proximity of the 5' fusion site to the minimal 35 bp promoter used in this study. The colored circles A and E represent the 4 base pair scars 5-GGAG-3 and 5-GCTT-3, respectively. Fluorescence measurements are shown in Molecules of Equivalent Fluorescein (MEFLs). Figure reprinted with permission from CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology, by S Iverson *et al.*, ACS Synthetic Biology, 2016 [38].

Chapter 3

Materials and Methods

3.1 Materials

3.1.1 Media

The media used for growing the bacteria strains in this study are shown in Table 3.1 below. All media were autoclaved (121 C for 15 minutes) or filter-sterilized (0.22 μm filter) prior to use.

Table 3.1: Bacterial growth media used in this study

Medium	Vendor	Used for
Luria Bertani (LB) Broth (Lennox)	Sigma-Aldrich	General growth in liquid media
LB Broth with Agar (Lennox)	Sigma-Aldrich	General growth in solid media
SOC Broth	Sigma-Aldrich	Post-transformation recovery
L-Arabinose	Sigma-Aldrich	Small molecule inducer

3.1.2 Bacterial strains

All transformation and flow cytometry assays were performed in Alpha Select Gold Efficiency *E. coli* cells (Bioline USA Inc., Tauton, MA, USA) (Genotype: F- *deoR endA1 recA1 relA1 gyrA96 hsdR17*(rk-, mk+) *supE44 thi-1 phoA* Δ (lacZYA-argF)U169 ϕ 80*lacZ* Δ M15 λ -). α -Select Gold cells are directly comparable to DH5 α cells. They

contain a *lacZ* marker that provides α -complementation of the β -galactosidase gene for blue/white color screening.

Bacterial strains were streaked for single colonies from glycerol stocks onto solid agar media and grown overnight at 37°. Liquid cultures (see section 3.3) were inoculated with single bacterial colonies and grown overnight at 37° with shaking (300 rpm)

3.1.3 Antibiotics

The final (1x) concentrations of antibiotics used for selective growth of *E. coli* in both solid and liquid media in this study are listed in Table 3.2 below.

Table 3.2: **Antibiotics used in this study.** Relevant antibiotics were added to all bacterial growth media in 1x concentration.

Antibiotic	Final Concentration	Vendor
Ampicillin	100 $\mu\text{g}/\text{mL}$	Zymo Research
Kanamycin	35 $\mu\text{g}/\text{mL}$	Zymo Research
Chloramphenicol	25 $\mu\text{g}/\text{mL}$	Zymo Research

3.1.4 Kits

Commercially available kits as available from vendors listed below in Table 3.3 were used in this study as per manufacturer’s protocols.

3.1.5 Enzymes

The enzymes used in this study are described in Table 3.4. All enzymes were purchased either from New England BioLabs (Ipswich, MA, USA) or Promega Corp. (Madison, WI, USA).

Table 3.3: **Molecular kits used in this study.** Miniprep and maxiprep kits were used to purify DNA for sequence verification or for use in downstream DNA assembly reactions. PCR cleanup kits were used to remove PCR reagents and oligonucleotide primers from reactions before using in downstream DNA assembly steps.

Kit	Company	Used for
QIAprep Spin Miniprep Kit	Qiagen	Purification of plasmid DNA
GenCatch Plasmid DNA Mini-Prep Kit	Epoch Life Science	Purification of plasmid DNA
NucleoBond Xtra Midi Plus	Machery-Nagel	Purification of plasmid DNA
QIAquick PCR Purification Kit	Qiagen	Purification of PCR product
GenCatch PCR Cleanup Kit	Epoch Life Science	Purification of PCR product
QIAquick Gel Extraction Kit	Qiagen	Purification of DNA from agarose gel
GenCatch Gel Extraction Kit	Epoch Life Science	Purification of DNA from agarose gel

3.1.6 DNA parts and primers

The following basic DNA parts in 3.5 were taken from the CIDAR Inventory of Composable Elements (ICE) Registry [56] and available from Addgene [1] (# 1000000059) for use in the assembly of some of the higher level circuits in this study. Rationally insulated promoter parts and promoter parts insulated using our new custom DNA spacer generation and screening method were synthesized as gblocks from Integrated DNA Technologies (Coralville, Iowa, USA). Rationally and empirically (using new screening technique) insulated promoter parts that were used to create Level 1 MoClo circuits and Level 2 inverter devices are listed in Table 3.8 and Table 3.9, respectively.

All primers for sequencing and creating randomized insulated promoter libraries are listed in Table 3.11 and Table 3.10, respectively. Primers VF and/or VR were used to sequence all intermediate inverter MoClo Level 1 circuits and to run colony

Table 3.4: **Enzymes used in this study.** BsaI was used to construct Level 1 genetic circuits from basic (Level 0) parts. BbsI was used to construct Level 2 MoClo circuits from Level 1 circuits. Restriction digests were performed to verify genetic circuit size or to move a circuit into a different vector backbone. DpnI was used in the creation of insulated promoter libraries as described in section 3.2.3.1

Enzyme	Vendor	Application
BsaI	New England BioLabs	MoClo DNA Assembly
BbsI	New England BioLabs	MoClo DNA Assembly
T4 Ligase	Promega	MoClo DNA Assembly
T4 Ligase	New England BioLabs	Ligation
EcoRI	New England BioLabs	Restriction Digestion
XbaI	New England BioLabs	Restriction Digestion
SpeI	New England BioLabs	Restriction Digestion
PstI	New England BioLabs	Restriction Digestion
DpnI	New England BioLabs	Removing PCR template DNA

PCRs (Section 3.2.4). Full inverters were sequence-verified with sequencing primers designed to AraC, TetR, GFP and RFP CDSs as listed in Table 3.10.

3.2 Methods

3.2.1 DNA purification

Plasmid DNA were purified from target bacterial cells grown in a 5 mL overnight LB culture supplemented with the appropriate antibiotic using the QIAprep Spin Miniprep kit (Qiagen), GenCatch Plasmid DNA Mini-prep kit (Epoch Life Science) or NucleoBond Xtra Midi Plus kit (Macherey-Nagel) according to the manufacturers instructions. The plasmids were eluted in a final volume of 50 μ L of GenCatch Elution Buffer (10 mM Tris-Cl, pH 8.5) for mini-prep protocols and in 300 μ L GenCatch Elution Buffer for midi-prep protocols.

3.2.2 Modular Cloning (MoClo) DNA assembly

An updated and optimized protocol for the single pot, multi-part, Type IIS restriction endonuclease-based Modular Cloning (MoClo) DNA assembly method first described in 2011 [68] was used to generate all MoClo Level 1 monocistronic inverter/NOT-gate circuits in Chapters 4 and 6. All basic DNA parts (Table 3.5) were either obtained from the CIDAR-ICE registry [56, 30] or were purchased as gblocks from Integrated DNA Technologies (Madison, WI, USA) with the addition of DNA spacers and 4 bp MoClo-format overhangs (fusion sites) for assembly into MoClo Level 1 circuits. Level 1 MoClo circuits were assembled into Level 2 MoClo circuits (inverters) by standard MoClo assembly as described here [30, 38].

3.2.2.1 MoClo Level 1 part assembly from basic DNA parts in MoClo format

A transcriptional unit (promoter::RBS::CDS::terminator) or Level 1 MoClo part can be constructed from basic/Level 0 MoClo parts by combining equimolar amounts (4-40 fmol) of each basic part into a Level 1 destination vector in a one pot restriction digestion-ligation reaction using BsaI and T4 DNA ligase enzyme (3.4). Level 0 parts have overlapping fusion sites such that the 3' fusion site of the upstream part must match the 5' fusion site of the downstream part for accurate directional assembly. The destination vector fusion sites must match the 5' fusion site of the promoter and the 3' fusion site of the terminator respectively. Level 0 parts have either chloramphenicol or ampicillin resistance while Level 1 vectors have kanamycin resistance and use *lacZ α* blue-white selection. An overview of MoClo Level 1 part assembly is schematically shown in Figure 3.1

3.2.2.2 MoClo Level 2 part assembly from Level 1 MoClo parts

Level 2 MoClo parts (higher-level genetic devices comprising multiple MoClo monocistronic transcriptional units) can be made by combining equimolar amounts of Level 1 MoClo parts (individual transcriptional units) with an ampicillin-resistant Level 2 destination vector in a one pot, restriction digestion-ligation reaction using BbsI (also called BpiI) and T4 DNA ligase (Table 3.4). The 3'-flanking fusion site of the first transcriptional unit must overlap with the 5'-flanking fusion site of the second transcriptional unit etc., for accurate directional assembly. Level 2 MoClo reactions also use *lacZ α* blue-white selection. An overview of MoClo Level 1 part assembly is schematically shown in Figure 3.2

3.2.2.3 Modular Cloning (MoClo) Reaction Parameters and reaction protocol:

Each of 4 DNA parts (Level 0 parts for Level 1 MoClo reaction, and 2-4 Level 1 MoClo parts for Level 2 MoClo reaction) along with destination vector (DVL1 or DVL2 for Level 1 and Level 2 MoClo reactions, respectively) with fusion sites appropriate to insert parts (see Figures 3.1 and 3.2) were diluted to 10-20 fmol concentrations and added in equimolar amounts to the reaction. DNA parts can be purified plasmid DNA, PCR fragment or synthesized double-stranded DNA. DNA parts were combined with 1x T4 DNA ligase buffer (Promega), 5 units of BsaI (New England Biolabs) for Level 1 MoClo reaction and 10 units for Level 2 MoClo reaction, along with 120 units of T4 DNA ligase enzyme (New England BioLabs). The reaction was brought up to a final volume of 10 μ L. MoClo reactions were performed in an Eppendorf Mastercycler ep thermocycler (Eppendorf) using the following parameters: 25 cycles (37°C for 1.5 minutes, 16°C for 3 minutes), 50°C for

5 minutes and 80° for 10 minutes. Reactions were then held at either 4° or -20° until transformed.

3.2.3 Randomized spacer generation and screening

3.2.3.1 Randomized spacer generation

A 100% sequence-randomized, 36 bp insulated promoter library for a target MoClo promoter was generated by performing an inverse PCR reaction using divergent primers containing phosphorylated 18N (NNNNNNNNNNNNNNNNNNNN) 5' overhangs using either a constitutively expressed GFP expression cassette (J23100), an inducible GFP expression cassette (pBAD) or a repressible RFP expression cassette (pTet) as template. PCR was followed by a blunt-end ligation followed by transformation and plating on LB agar with appropriate antibiotics.

To enable subsequent screening of insulated promoters (see Figure 3.3), the target promoter was assembled into a GFP-expression cassette and the resulting transcriptional unit (TU) was used as the PCR template. The uninsulated transcriptional unit plasmid DNA was eliminated by overnight DpnI digestion followed by overnight blunt-end ligation as shown in Figure 3.3.

The forward primer was designed to include (from 5' to 3') a 100% degenerate 18N sequence followed by the promoter and further downstream sequence up to 44 bases. The reverse primer was designed to include (from 5' to 3') a 100% degenerate 18N sequence, followed by the MoClo fusion site immediately upstream of promoter and further upstream sequence up to 44 bases. The primers were ordered phosphorylated and PAGE-purified from the manufacturer (Integrated DNA Technologies, Iowa, USA). Eight 50 μ L PCR reactions were set up using High Fidelity (HF) Phusion polymerase (New England BioLabs, Ipswich, MA, USA). Reactions were set up as per

the manufacturer's recommended protocol. The PCR products were pooled together. Template plasmid DNA was eliminated from the PCR product by overnight DpnI digestion of PCR product at 37° followed by heat inactivation of enzyme at 80° for 20 minutes. DpnI digestion was carried out as outlined in Section 3.2.3.2. The digested PCR product was purified (Table 3.3) and held at at -20° until ligation.

3.2.3.2 Ligation and transformation

Blunt-end ligation was performed on the purified, digested and heat-inactivated PCR product at 16°C overnight in an Eppendorf Mastercycler ep thermocycler (Eppendorf). Ligated product was either held on ice or stored at -20° until ready to be transformed. The ligated PCR product contains the fully constructed insulated promoter cassette library. 5 μ L of ligated insulated promoter expression cassette library was transformed into 25 μ L of *E. coli* Alpha Select Gold Efficiency chemically competent *E. coli* cells (BioLine USA Inc., Tauton, MA, USA) as described in Section 3.2.5.

3.2.3.3 Randomized spacer screening

For tested promoters, spacer screening was performed using the upstream (5') MoClo fusion site that most affected promoter expression. Colonies from the insulated promoter expression cassette library were grown overnight in LB broth with appropriate antibiotics. Fluorescence expression for colonies were tested by flow cytometry for either GFP (promoters J23100 and pBAD) or RFP (promoter pTet). For constitutive promoter J23100, insulated J23100 samples with GFP expression levels identical to the J23100-GFP_A cassette were selected from the tested samples as candidates for DNA spacers in target genetic devices. For promoter pBAD, insulated pBAD-GFP

samples with GFP expression identical to that of pBAD-GFP_A in the induced state were selected as candidate DNA spacer sequences. For repressible promoter pTet, insulated pTet-RFP samples demonstrating RFP repression of >10x were selected as candidate spacer sequences.

3.2.4 Colony PCR

Fully assembled inverter devices were first checked for size by performing a colony PCR using standard PCR reagents and VF and VR primers (Table 3.10) to check the size of the cloned product after the MoClo Level 2 reaction (Section 3.2.2, Figures 3.2, 3.2.2). Taq 2x MasterMix (New England BioLabs, Ipswich, MA, USA) and 0.2 μ M each of primers VF and VR (Table 3.10). PCR reaction was performed with an initial denaturation step at 95°C for 30 seconds, followed by 25 cycles each of 95°C for 20 seconds, 61°C for 20 seconds (primer annealing) and 68°C for 1 minute per kilobase of DNA in reaction (primer extension). After cycling, a final extension of 7 minutes at 68°C was performed. Reactions were held at 4°C until ready to be checked on a gel. All inverters were checked for size by electrophoresis through a 1% agarose gel using a 2-log ladder (New England BioLabs, Ipswich, MA, USA) as molecular size markers. Inverters that were the correct size were sequenced using sequencing primers designed to AraC, TetR, GFP and RFP CDSs (Table 3.10).

3.2.5 Transformation and selection

For transformation of MoClo reactions, 3 μ L of the MoClo (Level 1 or Level2) reaction mixture was combined with minimal handling with 7-20 μ L of Alpha Select Gold Efficiency chemically competent *E. coli* cells (Biolone USA Inc., Tauton, MA, USA) that had been thawed on ice. For transformation of ligation mixture from blunt-end

ligation for randomized insulated cassette generation, 5 μL of the ligation mixture was combined with 25 μL of Alpha Select Gold Efficiency chemically competent *E. coli* cells (Bioline USA Inc., Tauton, MA, USA) that had been thawed on ice.

The mixture was allowed to incubate on ice for 2-15 minutes before being heat shocked in a 42° water bath for 30-45 seconds. The cells were recovered with 1 mL of SOC broth for 45-60 minutes at 37° shaking (300 rpm) before being plated on appropriate antibiotic selective LB agar plates supplemented with 80 μL of 20mg/ μL 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside (X-GAL) and 100 μL of 0.1M isopropyl- β -D-thiogalactopyranoside (IPTG) (Zymo Research Corp., Irvine, CA USA).

White colonies were selected from the overnight transformation plate and grown overnight at 37° shaking in LB broth medium with appropriate antibiotic. For MoClo reactions, plasmid DNA was purified using DNA purification using plasmid purification kits as per manufacturer's instructions (see Section 3.3) and sent for DNA sequencing. For randomized spacer generation, selected colonies were screened by flow cytometry (see Section 3.3.2).

3.2.6 DNA sequencing

For purified plasmid DNA sequencing, a 10 μL reaction was prepared containing 150-200 ng of purified plasmid DNA, 1.67 pmol/ μL sequencing primer (1 μL primer from a 10 μM stock) and deionized water to 10 μL . The spacer library was sequenced by submitting unpurified colony PCR samples and 10 μM primer stock. All sequencing reactions were submitted to QuintaraBio for Sanger sequencing. Sequence analysis of trace files were performed using Benchling [60, 15].

3.3 In vivo assay methods

3.3.1 Growth of bacterial cells

Single bacterial colonies were inoculated into culture tubes with LB broth with appropriate antibiotics. Inoculated cultures were grown 14-17 hours to a density of approximately 34×10^9 cells per mL.

3.3.2 Fluorescence measurements by flow cytometry

All fluorescent expression devices were characterized using a BD LSRFortessa SORP flow cytometer. RFP fluorescence was measured using a solid-state Coherent Sapphire 561 nm laser at 100 mw strength with a PE-Texas Red 610/20 filter. GFP fluorescence was measured using a solid-state Coherent Sapphire 488 nm laser at 200 mw strength with a FITC 530/30 filter. LB agar (Sigma-Aldrich, St. Louis, MO, USA) plates containing the appropriate antibiotic were inoculated with Alpha-Select Gold Efficiency *E. coli* cells (Bioline) containing the confirmed plasmid construct/clone from insulated promoter cassette library of interest. Colonies were grown in 200 μ L LB broth (Sigma-Aldrich) with the appropriate antibiotic in sterile 96-well, deep well plates (BioExpress, Kaysville, UT, USA) in triplicate (for confirmed plasmid construct) or singly (for insulated promoter expression cassette library) for 14 hours at 37°C shaking at 300 rpm. Cells were then diluted 100-fold into 200 μ L of 1x sterile phosphate buffered saline (BioExpress) in 96-well round bottom plates before measurement using a high-throughput sampler (HTS).

3.3.3 Molecules of Equivalent Fluorescein (MEFL) conversion

Flow cytometry data was converted from arbitrary units to compensated MEFL (Molecules of Equivalent Fluorescein) using the TASBE characterization method

[9, 7]. An affine compensation matrix is computed from single color, dual color, and blank controls: RFP (red) alone (J23104:BCD2:E1010m:B0015 in the MoClo Level 1 destination vector DVL1_AE, abbreviated as pJ04B2Rm_AE), GFP (green) alone (J23104:BCD2: E0040m:B0015 in the MoClo Level 1 destination vector DVL1_AE, abbreviated as pJ04B2Gm_AE, the dual red-green color control (pJ04B2Rm:J04B2Gm) in the MoClo destination vector DVL2_AF, abbreviated as pJ04B2Rm:J04B2Gm_AF) and untransformed Alpha Select *E. coli* cells (Bioline, Tauton, MA, USA), respectively. FITC measurements (for GFP) are calibrated to MEFL using SpheroTech RCP-30-5-A beads [36].

Table 3.5: **Basic MoClo DNA parts and vectors used in this study to construct higher level genetic circuits.** All basic parts (except B0030m_BC) are archived and publicly available through Addgene [1] and the CIDAR-ICE DNA Parts Registry [56]. The following basic parts were used to construct higher level circuits and devices described in this study. Insulated promoters were derived from basic promoter parts listed below by inserting additional bases (12, 24 or 36 bp) to serve as DNA spacers between the promoter and its 5 flanking MoClo fusion site.

Basic DNA part name	Part Type	CIDAR-ICE ID	Addgene ID
J23100_AB	constitutive promoter	CIDAR-490	65980
J23100_EB	constitutive promoter	CIDAR-491	65981
J23100_FB	constitutive promoter	CIDAR-492	65982
J23100_GB	constitutive promoter	CIDAR-493	65983
pBAD_AB	inducible promoter	CIDAR-526	66016
pBAD_EB	inducible promoter	CIDAR-527	66017
pBAD_FB	inducible promoter	CIDAR-528	66018
pBAD_GB	inducible promoter	CIDAR-529	66019
pTet_AB	repressible promoter	CIDAR-518	66008
pTet_EB	repressible promoter	CIDAR-519	66009
pTet_FB	repressible promoter	CIDAR-520	66010
pTet_GB	repressible promoter	CIDAR-521	66011
B0030m_BC	RBS/BCD		74394
B0032m_BC	RBS/BCD	CIDAR-530	66020
B0034m1_BC	RBS/BCD	CIDAR-531	66022
BCD12_BC	RBS/BCD	CIDAR-533	66023
E0040m_CD (GFP)	CDS	CIDAR-542	66032
E1010m_CD (RFP)	CDS	CIDAR-543	66033
C0040_CD (TetR)	CDS	CIDAR-537	66027
C0080_CD (AraC)	CDS	CIDAR-539	66029
B0015_DE	Double terminator	CIDAR-545	66035
B0015_DF	Double terminator	CIDAR-546	66036
B0015_DG	Double terminator	CIDAR-547	66037
B0015_DH	Double terminator	CIDAR-548	66038
DVLK_AE	destination vector	CIDAR-577	66067
DVLK_EF	destination vector	CIDAR-578	66068
DVLK_FG	destination vector	CIDAR-579	66069
DVLK_GH	destination vector	CIDAR-580	66070
DVLA_AH	destination vector	CIDAR-553	66043
DVLA_AF	destination vector	CIDAR-551	66041
DVLA_AG	destination vector	CIDAR-552	66042

Table 3.6: **Promoter parts insulated with 12 bp rationally designed spacers.** Constitutive promoters J23100, J23101, J23104, inducible promoter pBAD and the repressible promoter pTet were insulated with 12 bp spacers. DNA spacers were placed between MoClo fusion site and immediately 5' of promoter. Sequence is read as MoClo 5 fusion site-Spacer-promoter-MoClo 3 fusion site. First four bases represent upstream MoClo fusion site. DNA spacer sequence is represented in capital letters. Promoter sequence follows spacer sequence. Final four bases represent the MoClo 3 fusion site immediately downstream of the promoter. The sequences listed below can also be found in the CIDAR Benchling parts collections.

Rationally Insulated Promoters	Sequence
12INS1-J23100_E	gcttTCCCGAGTAAGTttgacggctag ctcagtcctaggtacagtgctagctact
12INS2-J23100_E	gcttCATATCTCCAAAttgacggctag ctcagtcctaggtacagtgctagctact
12INS1-J23101_E	gcttTCCCGAGTAAGTtttacagctag ctcagtcctaggtattatgctagctact
12INS2-J23101_E	gcttCATATCTCCAAAtttacagctag ctcagtcctaggtattatgctagctact
12INS1-J23104_E	gcttTCCCGAGTAAGTttgacagctag ctcagtcctaggtattgtgctagctact
12INS2-J23104_E	gcttCATATCTCCAAAttgacagctag ctcagtcctaggtattgtgctagctact
12INS1-pTet_E	gcttTCCCGAGTAAGTtcctatcagtg atagagattgacatccctatcagtgatagagatac tgagcactact
12INS2-pTet_E	gcttCATATCTCCAAAtcctatcagtg atagagattgacatccctatcagtgatagagatac tgagcactact

Table 3.7: **Promoter parts insulated with 24 bp spacers.** Constitutive promoters J23100, J23100, J23104, inducible promoter pBAD and the repressible promoter pTet were insulated with 24 bp rationally designed spacers. DNA spacers were placed between MoClo fusion site and immediately 5' of promoter. Sequence is read as MoClo 5 fusion site-Spacer-promoter-MoClo 3 fusion site. First four bases represent upstream MoClo fusion site. DNA spacer sequence is represented in capital letters. Promoter sequence follows spacer sequence. Final four bases represent the MoClo 3 fusion site immediately downstream of the promoter. The sequences listed below can also be found in the CIDAR Benchling parts collections.

Rationally Insulated Promoters	Sequence
24INS1-J23100_E	gcttCCATTGAGCCAGCTTCAGGGTTCA ttgacggctagctcagtcctaggtacagtgctagctact
24INS1-J23100_E	gcttGAGGTCCGTGCACCGTCAGAAGCC ttgacggctagctcagtcctaggtacagtgctagctact
24INS1-J23101_E	gcttCCATTGAGCCAgcttCAGGGTTCA tttacagctagctcagtcctaggtattatgctagctact
24INS1-J23101_E	gcttGAGGTCCGTGCACCGTCAGAAGCC tttacagctagctcagtcctaggtattatgctagctact
24INS1-J23104_E	gcttCCATTGAGCCAgcttCAGGGTTCA ttgacagctagctcagtcctaggtattgtgctagctact
24INS1-J23104_E	gcttGAGGTCCGTGCACCGTCAGAAGCC ttgacagctagctcagtcctaggtattgtgctagctact
24INS1-R0040_E	gcttCCATTGAGCCAgcttCAGGGTTCA tcctatcagtgatagagattgacatccctatcagtgatag agatactgagcactact
24INS1-R0040_E	gcttGAGGTCCGTGCACCGTCAGAAGCC tcctatcagtgatagagattgacatccctatcagtgatagag atactgagcactact

Table 3.8: **Promoter parts insulated with 36nt spacers.** 36nt DNA spacers were placed between 5' boundary and upstream MoClo fusion site of promoters J23100, J23100, J23104, pBAD and pTet. First and final 4 bp represent 5' and 3' MoClo fusion sites. DNA spacer sequence is in capital letters. Promoter sequence follows spacer sequence. Final four bases represent the MoClo 3 fusion site immediately downstream of the promoter. Table continued on following page.

Insulated Promoter	Insulated Promoter Sequence
36INS1-J23100_E	gcttCCTGCAGTATTCATTTTCAGC TTACGGAAGGTAGATttgacggetagc tcagtcctaggtacagtgctagctact
36INS2-J23100_E	gcttAGCAGT TACTCAAGTGGTCA GGAGGCAGACGCAGCGttgacggct agctcagtcctaggtacagtgctagctact
36INS1-J23100_A	ggagCCTGCAGTATTCATTTTCAG CTTACGGAAGGTAGATttgacggcta gctcagtcctaggtacagtgctagctact
36INS2-J23100_A	ggagAGCAGT TACTCAAGTGGTC AGGAGGCAGACGCAGCGttgacgg ctagctcagtcctaggtacagtgctagctact
36INS3-J23100_A	ggagCCTTTTCACGAGTCGCGTAC TCTCCTAACTATCAAGttgacggctag ctcagtcctaggtacagtgctagctact
36INS4-J23100_A	ggagTTAATCAGTGATAAGTTAAGA TGCAACCAGAAGGCAttgacggctagct cagtcctaggtacagtgctagctact
36INS5-J23100_A	ggagACCCGTACCGGAGCTCAATAA CGGTGCGGCGTTTGGttgacggctagctc agtctaggtacagtgctagctact
36INS6-J23100_A	ggagAAGCGTTACAGCTCTTGTCCG TCCTAGGCAGCACTTttgacggctagctc agtctaggtacagtgctagctact
36INS7-J23100_A	ggagAGGGACGACCCGCGCACAGA ATGTGTCGGGGTAACAttgacggctagc tcagtcctaggtacagtgctagctact

Table 3.8: **Promoter parts insulated with 36nt spacers.** (Table continued from previous page.)

Insulated Promoter	Insulated Promoter Sequence
36INS8-J23100_A	ggagGGGTACAAAATTA AAAACAC AGTGATGGAACCGTCCttgacggctag ctcagtcctaggtacagtgctagctact
36INS9-J23100_A	ggagCATCTAGTTGCGGAAAGTGTG TCACGAGTATCACGTttgacggctagctca gtcctaggtacagtgctagctact
36INS10-J23100_A	ggagTAGGGGGCAAACATAACTCTC GTTCCGGGGAATGGCttgacggctagctc agtccctaggtacagtgctagctact
36INS1-J23101_E	gcttCCTGCAGTATTCATTTTCAGCTT ACGGAAGGTAGATtttacagctagctcagtc ctaggtattatgctagctact
36INS2-J23101_E	gcttAGCAGTTACTCAAGTGGTCAGG AGGCAGACGCAGCGtttacagctagctcagtc cctaggtattatgctagctact
36INS1-J23104_E	gcttCCTGCAGTATTCATTTTCAGCTTA CGGAAGGTAGATttgacagctagctcagtcct aggtattgtgctagctact
36INS2-J23104_E	gcttAGCAGTTACTCAAGTGGTCAGGA GGCAGACGCAGCGttgacagctagctcagtc taggtattgtgctagctact
36INS1-pTet_E	gcttCCTGCAGTATTCATTTTCAGCTTAC GGAAGGTAGATtcctatcagtgatagagattgac atccctatcagtgatagagatactgagcactact
36INS2-pTet_E	gcttAGCAGTTACTCAAGTGGTCAGGAG GCAGACGCAGCGtcctatcagtgatagagattgac atccctatcagtgatagagatactgagcactact

Table 3.9: **List of 36nt insulated promoters used in this study.** The following promoters were created using DNA spacers selected using my screening methodology as described in Chapter 3, Section 3.3 and are available through Addgene [1]. Insulated promoter parts will also be submitted to the Registry of Biological Parts [35].

Insulated promoters from screen	Length of DNA spacer	Addgene ID
36INS1-J23100_A	36 bp	78666
36INS1-J23100_E	36 bp	78667
36INS1-J23100_F	36 bp	78668
36INS1-J23100_G	36 bp	78669
A05-36INS-pBAD_A	36 bp	78670
A05-36INS-pBAD_E	36 bp	78671
A05-36INS-pBAD_F	36 bp	78672
A05-36INS-pBAD_G	36 bp	78673
D06-36INS-pBAD_A	36 bp	78674
D06-36INS-pBAD_E	36 bp	78675
D06-36INS-pBAD_F	36 bp	78676
D06-36INS-pBAD_G	36 bp	78677
C12-36INS-pTet_A	36 bp	78678
C12-36INS-pTet_E	36 bp	78679
C12-36INS-pTet_F	36 bp	78680
C12-36INS-pTet_G	36 bp	78681

Table 3.10: List of primers used for sequencing.

Primer Name	Sequence
VF	tgccacctgacgtctaagaa
VR	attaccgcctttgagtgagc
AraC-seq-for	attcggagctgctggcgataaatct
AraC-seq-rev	tgatcgctgatgtactgacaagcct
TetR-seq-for	acgctaaaagtttagatgtgctttact
TetR-seq-rev	tgagtgcataataatgcattctctagt
GFP-seq-for	acgtgctgaagtcaagttgaaggt
GFP-seq-rev	tgccatgatgtatacattgtgtgagt
RFP-seq-for	aggtttcaaatgggaacgtgttatgaa
RFP-seq-rev	ggaagttggtaccacgcagtttaact

Table 3.11: List of primers used to generate insulated promoter libraries.

Primer Name	Sequence	Function
36N-pJ00B12Gm -For	NNNNNNNNNNNNNNNNNNNNNN ttgacggctagctcagtcctaggtacagtgct agcTACTgggccc	36N-J23100_A library creation
36N-pJ00B12Gm _A-Rev	NNNNNNNNNNNNNNNNNNNNNN CTCCatgtcttccactagtctctagaagcg gccgccaattccag	36N-J23100_E library creation
36N-pJ00B12Gm _E-Rev	NNNNNNNNNNNNNNNNNNNNNN AAGCatgtcttccactagtctctagaagc ggccgccaattccag	36N-J23100_H library creation
36N-pJ00B12Gm _H-Rev	NNNNNNNNNNNNNNNNNNNNNN TAGTatgtcttccactagtctctagaagcg gccgccaattccag	36N-J23100_K library creation
36N-pJ00B12Gm _K-Rev	NNNNNNNNNNNNNNNNNNNNNN GACTatgtcttccactagtctctagaagcg gccgccaattccag	36N-pBAD_ library creation-1
36N-pBAD-For	NNNNNNNNNNNNNNNNNNNNNN acattgattatttgcacggcgtcacactttgct atgcatagca	36N-pBAD_ library creation-2
36N-DVL2_A-Rev	NNNNNNNNNNNNNNNNNNNNNN CTCCtgagaccactagtctctagaagcg gccgccaattccaga	36N library common reverse primer
36N-pTet-For	NNNNNNNNNNNNNNNNNNNNNN tcctatcagtgatagagattgacatccctatc agtgatagaga	36N-pTet_ library creation

Figure 3.1: **Construction of a Level 1 MoClo part.** Four basic DNA parts (promoter, RBS/5'UTR, CDS and terminator) are each shown in green vectors to denote the chloramphenicol resistance cassette and each has two 4 bp non-palindromic fusion sites flanking the DNA part (colored circles). The fifth vector is the Destination Vector (DV) for the final construct, which is orange to denote the kanamycin resistance cassette and contains the *lacZ* gene fragment for blue-white screening. The BsaI sites (green text) are shown in the DNA part vectors and DV with the direction they cut (green arrows). The 4 bp fusion sites (black text in colored circles) shown in the five top plasmids indicate the fusion sites that will remain with either the DNA parts or the DV after BsaI digestion. (caption continued on next page.)

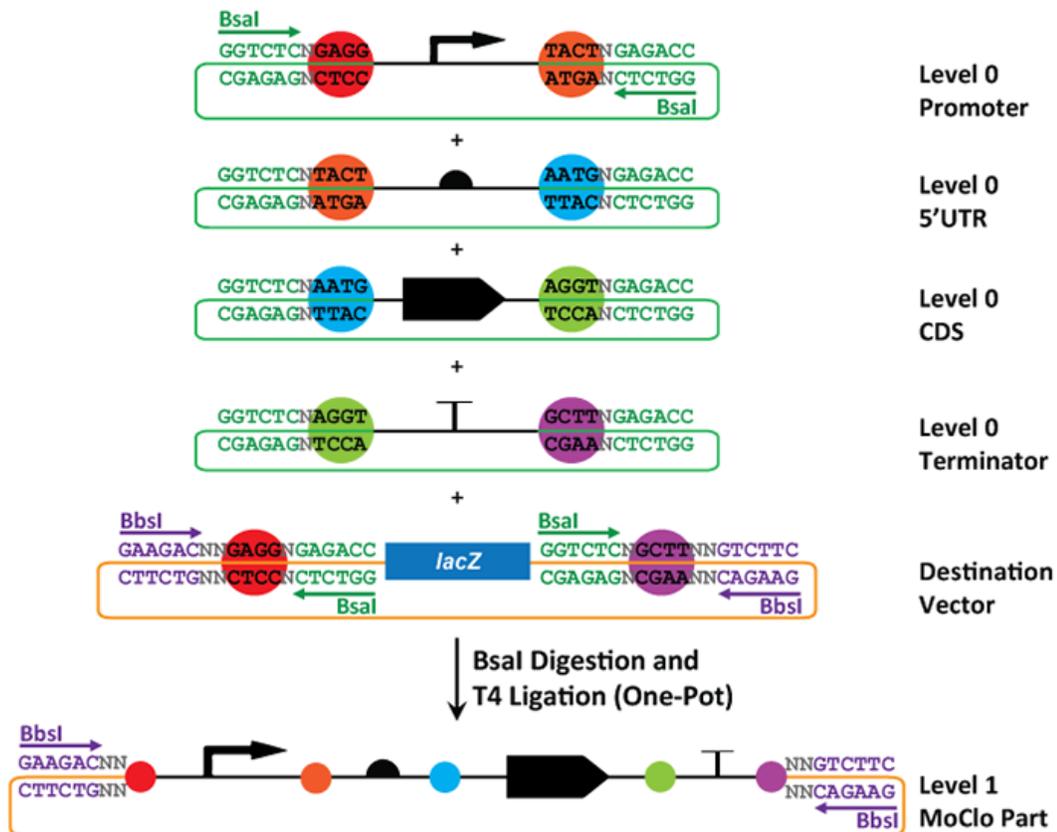


Figure 3.1: **Construction of a Level 1 MoClo part.** (Continued from previous page.) By allowing matching fusion sites between the 3' end of the first part and the 5' end of the second part, and so on, this ensures that the parts assemble in the desired assembly order of promoter-RBS/5'-UTR/CDS-terminator. The DV fusion sites match the fusion sites on the 5' end of the promoter and the 3' end of the terminator to allow the properly assembled transcriptional unit to be cloned into the vector in the correct orientation. The assembled transcriptional unit replaces the *lacZ* fragment and colonies containing kanamycin-resistant plasmids with assembled transcriptional unit grow on antibiotic media with IPTG and X-GAL as white colonies, thus enabling blue-white screening along with antibiotic selection. Figure reproduced with permission from Type IIS Assembly for Bacterial Transcriptional Units: A Standardized Assembly Method for Building Bacterial Transcriptional Units Using the Type IIS Restriction Enzymes BsaI and BbsI by Traci Haddock *et al*, 2015 from BioBricks Foundation. [30]

Figure 3.2: Construction of a Level 2 MoClo part. Four MoClo Level 1 parts (see Figure 3.1) are each shown in orange vectors to denote the kanamycin resistance cassette and each has two 4 bp non-palindromic fusion sites (colored circles) flanking all of the DNA parts. The fifth vector is the Destination Vector (DV) for the final construct, which is blue to denote the ampicillin resistance cassette and contains a *LacZ* fragment for blue-white screening. The *BbsI* sites (purple text) are shown in the DNA part vectors and the DV with the direction they cut (purple arrows). By allowing matching fusion sites between the 3' end of T1 and the 5' end of P2, and so on, the final assembled device maintains the transcriptional units in the desired order. The assembled multi-TU device replaces the *lacZ* fragment and colonies containing kanamycin-resistant plasmids with assembled transcriptional unit grow on antibiotic media with IPTG and X-GAL (Section 3.2.5 as white colonies, thus enabling blue-white screening along with antibiotic selection. Figure reproduced with permission from Type IIS Assembly for Bacterial Transcriptional Units: A Standardized Assembly Method for Building Bacterial Transcriptional Units Using the Type IIS Restriction Enzymes *BsaI* and *BbsI* by Traci Haddock *et al*, 2015 from BioBricks Foundation. [30]

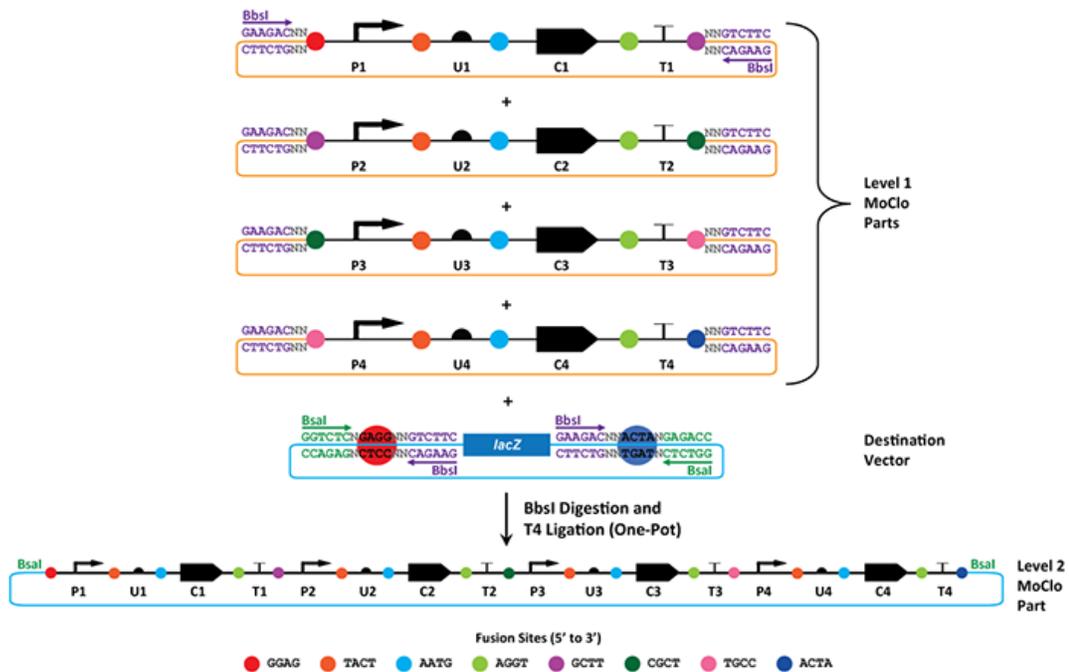
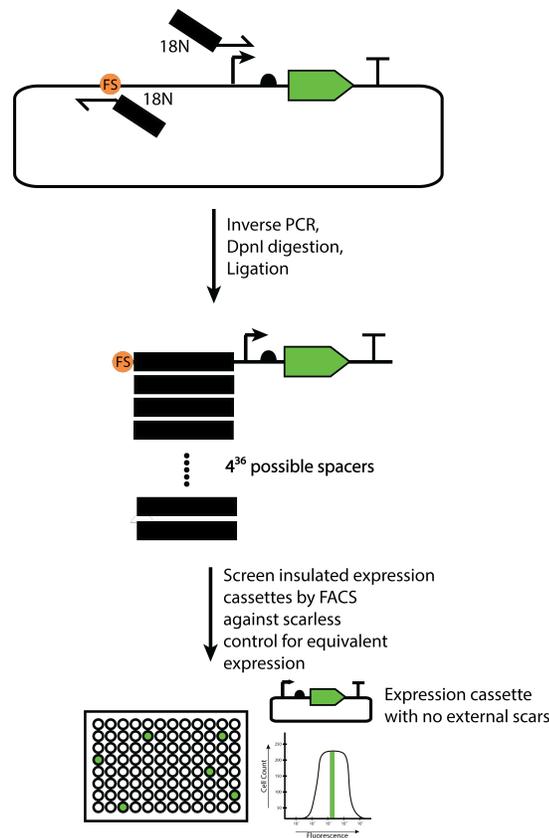


Figure 3.3: Divergent forward and reverse primers with 18N (NNNNNNNNNNNNNNNNNNNN) 5' primer extensions were used in an inverse PCR reaction using a plasmid containing the target genetic circuit as PCR template. The forward primer is designed to anneal to the promoter that is to be insulated. The reverse primer is designed to anneal to the MoClo fusion site immediately upstream of the promoter. The PCR reaction will result in a library of linearized plasmid DNA strands with the target promoter insulated. Overnight DpnI digestion removes residual template DNA. The linear plasmid PCR products are circularized in an overnight blunt-end ligation and then transformed and plated onto LB agar with appropriate antibiotics. Colonies from the library are grown overnight in liquid media, diluted before taking fluorescence measurements using a flow cytometer. The fluorescence expression of a reference device is used to screen the library of insulated expression cassettes for ones with the precise level of desired fluorescence expression. Plasmid DNA from the selected insulated cassettes are purified using commercial DNA purification kits (Table 3.3) and then sequenced. The DNA spacer sequence thus obtained is used to design new promoter parts for the final target device.



Chapter 4

DNA Spacers: Preliminary Work

4.1 Introduction

I begin this chapter with my initial exploration of component transcription unit order on the overall expression of a complex device. My test circuit in this experiment and in all subsequent component order investigations is a 4 transcription unit (TU) TetR inverter (NOT-gate) inducible by L-arabinose. In order to facilitate the measurement of the effect of component transcription unit order on device expression, I stripped my inverter circuit of all cis-regulation, decoupled the expression of AraC from its native pC promoter and put it under the control of the strong constitutive minimal promoter J23100. In this design, all 24 permutations of the four transcription units are valid inverter designs and should – in theory – have an identical expression profile (See Figure 4.1). My inverter circuit is induced through double inversion: the repressor AraC is constitutively expressed, which represses promoter pBAD. This AraC represses the expression of TetR and GFP from both instances of the pBAD promoter. Arabinose induction is achieved by growing the *E. coli* containing my inverter overnight (14 hours) in the presence of 1 mM L-arabinose. GFP and RFP measurements are taken in only the ON and OFF states of the inverter. Upon induction, the arabinose sugar binds to AraC and the pBAD promoter is induced resulting in the production of TetR and GFP proteins. TetR represses promoter

pTet and RFP expression is inhibited.

Genetic circuits are by convention assembled in the order in which the genetic logic is executed, probably because such a design is intuitive to understand and probably also because it is quicker to assemble this way and intuitively easy to follow such a device. Therefore, if a genetic device does not contain cis-regulation through DNA looping or similar mechanisms, there are no design constraints that necessitate that the modular genetic components be arranged in the conventional order. In my inverter circuit, each transcription unit is a modular component. Therefore, changing the order of the transcription units should yield devices of equivalent expression. However, fluctuations in the DNA sequences at the junctions of genetic parts or larger modular components are known to cause unpredictable changes in circuit expression [11, 13]. I wanted to assess the severity of such context effects.

I assembled the transcription units (TUs) of a 4 TU inverter (NOT-gate) in the 23 other possible tandem permutations of the four transcription units that my TetR inverter were composed of. All transcription units were still arranged in tandem to each other on the same strand of DNA. The assumption that I was testing was that all 24 permutations of my inverter circuit would nominally display basic inverter behavior: the expression of the output, RFP, would go down upon induction, and the expression of GFP, which is an indirect measure of the TetR (L-arabinose is added to saturation) would increase upon induction.

I assumed the alternate configurations would be equivalent in its expression profile to the device assembled in the conventional order. The 24 configurations of my single inverter demonstrated very different expression profiles, and it appeared that I saw a pattern of high GFP expression when pBAD-GFP was flanked by the scar “A” (5'-GGAG) immediately upstream of the pBAD promoter (Figure 4.2). In the most

stark instance, one inverter device showed a 22.5x increase in GFP expression change on induction compared to the conventional design where the GFP expression did not change at all. Later in Chapter 6, I show improvement in inverter expression profile as well as the improvement in the GFP and RFP expression fold changes on induction (Chapter 6).

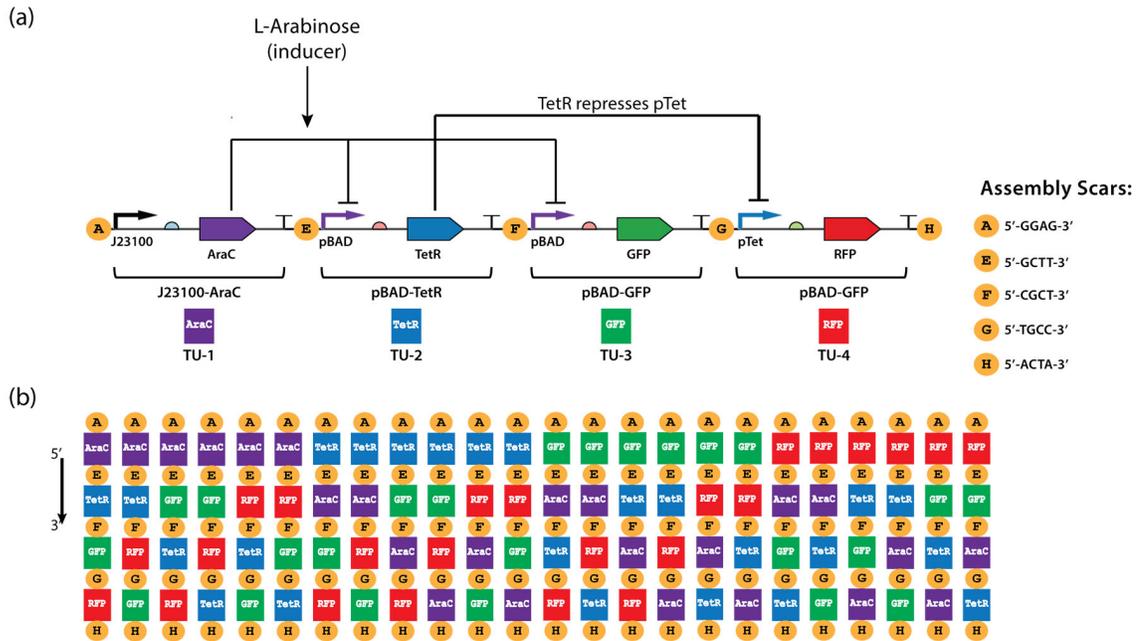


Figure 4.1: Bacterial transcriptional inverter (NOT-gate) and experimental design: (a) The experiments in this study were performed using a bacterial, monocistronic, transcriptional TetR-inverter composed of 4 tandem Transcriptional Units (TUs) assembled by the MoClo DNA assembly method. Circles between transcriptional units represent the 4 bp scars left behind 5' and 3' of each TU in the final construct post-assembly. The inverter is inducible by L-Arabinose. Native arabinose regulation was decoupled for greater control over device performance. All regulation is in trans. In the uninduced (OFF) state, AraC represses pBAD resulting in minimal TetR and GFP expression. In the absence of TetR, RFP is expressed. On induction (ON) with L-arabinose, TetR and GFP are expressed. TetR represses the promoter pTet, resulting in inhibition of RFP signal. (b) Keeping the orientation of the TUs unaltered, the order of TUs within the inverter are shuffled to produce 24 (4 x 3 x 2 x 1) different configurations of the same inverter. All permutations are done at the level of TU only.

When I started this work, I did not make any distinction between the order of the transcriptional units within the tested inverter and the order of 4 bp MoClo assembly scars at the 5' boundary of each promoter within my inverter that are left behind after the cloning process. However, as I progressed through my research, it became clear that the 4 bp scar at the 5' boundary of a promoter had a significant impact on promoter expression. I thus focused my research efforts on eliminating the variations that were caused by DNA fluctuations at the 5' junctions of promoters. While in my work, the DNA bases at the 5' boundary of a promoter happen to be a cloning scar that could have been avoided by the use of the Gibson DNA assembly method [26] or DNA synthesis, composing promoter parts into devices makes variations at the junctions of promoters unavoidable.

Finding no clear criteria for either DNA spacer length or base composition characteristics in the literature, I chose to begin with randomized DNA sequences ensuring that the candidate spacer sequences did not form any stable secondary structures that could possibly occlude transcription, or themselves have any transcription activating or enhancing properties. By chance the selected spacers fell within the normal base composition of the neutral DNA sequences found between *E. coli* operons that do not contain any known active gene regulatory elements [48]. I do not make any claims on testing a statistically significant number of DNA spacers in my preliminary work. Following up on the partial success of my preliminary work, I first expanded my DNA spacer set to avoid having to reuse a spacer twice within a single genetic device (See Figure 4.12). I recreated my 24 inverter permutations with promoters insulated with rationally designed DNA spacers, but found them to be unreliable (See Figure 4.13).

These failures led me to believe that designing DNA spacers rationally may not

be possible. This led to the development of the DNA spacer generation and high-throughput screening protocol. Using this technique I screened samples from libraries of insulated promoter circuits using the promoters from my test inverter (J23100, pBAD, pTet). These results are presented in Figures 5.2, 5.7 and 5.8 in Chapter 5.

4.2 Problem Statement: Synthetic genetic device performance is dependent on the order of intermediate transcription units

I assembled all 24 permutations possible from the reorganization of the transcription units of my 4 transcription unit inverter described in Figure 4.1 above using the MoClo DNA assembly method. The inverters were grown overnight in LB Broth as described in Chapter 3 and then grown for a second night in the presence (ON state) or absence (OFF state) or 1 mM L-arabinose. The inverters were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green and red fluorescence was measured by flow cytometry. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

I assessed the effect of changes in the local DNA sequence context of component modules on overall device performance. Local DNA context of component TUs is represented by the 4 bp MoClo assembly scar flanking the 5' boundary of each TU promoter. The 24 theoretically equivalent inverters varied greatly from configuration to configuration in their overall performance profiles. 8 out of 24 devices show almost no change in either their GFP and RFP expression levels upon induction. While GFP expression levels either change little or increase in all devices, in 8 of the

24 devices, the output (RFP) expression levels increase rather than decrease upon arabinose-induced TetR repression of the pTet promoter driving RFP expression. Furthermore, we observed a clear pattern of high GFP expression, but varying levels of RFP repression in all devices where the pBAD-GFP TU was flanked on the 5' end by the MoClo assembly scar "A" (pBAD-GFP_A).

4.3 Promoter Architecture for uninsulated promoters J23100, pBAD and pTet

The constitutive promoter J23100, the inducible promoter pBAD, and the repressible promoter pTet, used to design the 24 equivalent inverters (Figure 4.1 and Figure 4.2) were Modular Cloning (MoClo)-format versions of the J23100, pBAD and pTet promoters from the Registry of Standard Biological Parts [35]. No DNA sequence from the original promoter parts were deleted or rearranged in constructing the inverter circuits used in this study.

The promoter architecture for promoters J23100, pBAD and pTet without the 5' and 3' MoClo fusion sites (scars) are shown below in Figure 4.3, Figure 4.4 and, Figure 4.6, respectively. For each promoter, the first nucleotide marks the 5' boundary of the promoter and the last nucleotide marks the 3' boundary of the promoter. The 5' and 3' MoClo fusion sites begin at the 5' and 3' boundaries of each promoter.

4.3.1 Uninsulated J23100 promoter architecture

The constitutive promoter J23100 is a strong, minimal synthetic promoter that is part of a family of synthetic minimal constitutive promoters created by combinatorially varying the consensus sequences of a synthetic constitutive promoter created by combining the canonical RNA polymerase (RNAP) sigma subunit binding consensus

sequence centered at -35 and -10 relative to the transcription start site separated by a 17 base pair spacer. Promoter J23100 is a strong promoter because its expression level is closest to wild-type expression: its -35 sequence (5'-TTGACG-3') varies from the canonical consensus sequence (5'-TTGACA-3') by just one base and its -10 sequence (5'-TACAGT-3') varies from the canonical consensus sequence (5'-TATAAT-3') by two bases (see Figure 4.7).

4.3.2 Uninsulated pBAD promoter architecture

The inducible promoter pBAD used in this study is the MoClo version of the promoter I13453 from the Registry of Standard Biological Parts [35]. The synthetic promoter pBAD (see Figure 4.4) is 130 base pairs long and contains operator sequences AraI₁ and AraI₂ but not operator AraO₂, in addition to the -35 and -10 RNAP sigma subunit binding sites.

In the wild-type pBAD promoter, a dimer of AraC binds to the operators I₁ and O₂ in the repressing state, resulting in looping of the promoter over itself. Upon addition of the inducer, L-arabinose, AraC undergoes a conformational change making long-distance looping impossible and thus restricting AraC to binding proximal operator sites araI₁ and araI₂. The disappearance of the DNA loop and the occupation of the I₂ site lead to pBAD activation [52]. Figure 4.5 shows the mechanism of promoter pBAD repression and activation by the repressor, AraC, in the absence and presence of Arabinose, respectively.

The synthetic promoter (see Figure 4.4) is shorter (130 base pairs (synthetic) vs 284 base pairs(wild-type)) and is missing the wild-type operator araO₂ site that binds the C-terminus of one molecule of the AraC repressor dimer in the uninduced (-Ara) state. The synthetic promoter does not loop around in the uninduced state. The C-

terminus of one AraC monomer remains unbound in the synthetic pBAD promoter, which may possibly explain its leaky expression compared to the wild-type pBAD promoter.

Interestingly, the pBAD promoter is neither <100 base pairs in length, nor is there a protein binding site immediately downstream of the promoter 5' boundary. Yet, promoter pBAD is affected significantly by one (Scar "F", 5'-CGCT-3') of the four scars introduced upstream of its 5' boundary as a result of DNA assembly using the MoClo DNA assembly method (See Figure 4.8). It is possible that the scar "F" introduces long-range/off-target effects that are affecting promoter expression.

4.3.3 Uninsulated pTet promoter architecture

Promoter pTet (Figure 4.6) is a 54 base pair, regulatable promoter that binds, in addition to RNA polymerase enzyme, its cognate helix-turn-helix (HTH) repressor, TetR. Promoter pTet is constitutively expressed unless its repressor protein TetR is bound to the TetR-O₁ and TetR-O₂ sites. The pTet promoter used in the assembly of the inverters in this study is the MoClo-format version of the R0040 pTet promoter part from the Registry of Standard Biological Parts [35]. In the absence of the repressor, TetR, RNAP binds the -35 and -10 sequences. When TetR is present, it occupies the TetR-O₁ and TetR-O₂ sites, blocking RNAP access to the -35 and -10 sequences. No base pairs were deleted from the original Registry promoter part. The 5' MoClo scar gets added immediately upstream of the 5' boundary of the pTet promoter part. The TetR-O₁ ends at the 5' boundary of the promoter and is not buffered by any additional DNA sequence.

4.4 Characterization of promoter expression in isolation

4.4.1 Inverter component TU promoters are affected by 4 bp MoClo fusion sites immediately upstream of promoter

It has recently been reported that minimal constitutive promoters are affected by local DNA sequence context of as few as 4 bp of DNA immediately 5' of the promoter [38]. However, the study only investigated two MoClo fusion sites, fusion site "A" (5'- GGAG) compared to fusion site "E" (5'- GCTT); our 4-TU inverters contained in addition to fusion sites "A" and "E", fusion sites "F" (5'- CGCT) and "G" (5'- TGCC) upstream of promoter elements. Furthermore, the study did not investigate the effect of upstream DNA sequence on the expression of regulatable promoters like the inducible pBAD and the repressible pTet promoters used to build the 24 inverter sets. To assess whether the upstream 4 bp MoClo fusion sites could be responsible for the observed order dependent expression of the devices in the 24 inverter set, I tested each promoter-fusion site combination from my test inverter in isolation.

4.4.1.1 J23100_E and J23100_F expression is >10.5x lower than J23100_A expression

To test the effect of 4 bp MoClo fusion site 5' of promoter J23100, I built four individual GFP expression cassettes with the four 4 bp fusion sites, "A", "E", "F" and "G" upstream of promoter J23100 and measured their expression by flow cytometry. Circuits were grown overnight in LB Broth as described in Chapter 3. Circuits were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green fluorescence measured. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9]. J23100_A (8.27×10^4 MEFLs) and J23100_G ($7.95 \times$

10^4 MEFLs) expression varied minimally. However, J23100_E (6.63×10^3 MEFLs) and J23100_F (7.54×10^3 MEFLs) expression were observed to be 12.5x and 11.0x lower than J23100_A respectively.

The expression of the J23100-AraC_X (where X is either fusion site “A”, “E”, “F” or “G”) TU determines the amount of AraC protein in our inverters, which binds with the inducer L-arabinose to induce the expression of TetR to repress output RFP expression when the inverter circuit is induced. Lesser AraC between devices (dictated by the upstream fusion site based on the position of the J23100-AraC TU in the inverter) will likely impact inverter performance. It may be possible to regularize AraC levels between different inverter configurations in the 24 inverter set by insulating the J23100 promoter against its upstream DNA sequence context.

4.4.1.2 pBAD_F expression is 6.50x lower than pBAD_A expression

Regulatable promoters are considered less likely to be affected by fluctuations in the DNA sequence immediately upstream of the promoter [62, 22]. However, to our knowledge, this has not been investigated.

To test the effect of the DNA sequence context of a 4 bp MoClo fusion site 5' of inducible promoter pBAD, we built 4 individual GFP expression cassettes with the four 4 bp fusion sites, “A”, “E”, “F” and “G” upstream of promoter pBAD and measured their expression by flow cytometry (see Figure 4.8). Circuits were first grown in LB broth with appropriate antibiotics as described in Chapter 3, Section 3.3. Circuits were then diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green fluorescence measured. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9]. pBAD_G (6.76×10^3 MEFLs; 1.44x

higher than pBAD_A expression) had the highest expression, followed by pBAD_A (3.71×10^3 MEFLs) and pBAD_E (2.69×10^3 MEFLs, lower than pBAD_A by 1.38x). However, pBAD_F expression (5.75×10^2 MEFLs, 6.46x lower than pBAD_A) was significantly lower than pBAD_A and could be partly responsible for the observed order dependence of the 24 inverter set.

4.4.1.3 pTet promoter expression is unaffected by its upstream MoClo fusion site

To test the effect of the DNA sequence context of a 4 bp MoClo fusion site 5' of the repressible promoter pTet, we built 4 individual RFP expression cassettes with the four 4 bp fusion sites, "A", "E", "F" and "G" upstream of pTet and measured their expression by flow cytometry (see Figure 4.9). Circuits were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green fluorescence measured. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

Unlike the constitutive minimal promoter, J23100, and the inducible promoter pBAD, the pTet promoter was found to be robust and expression from it is not affected by the base composition of the 4 bp MoClo fusion site immediately upstream of the promoter in isolation. Insulating pTet in the 24 inverter sets appeared to not be essential to eliminating upstream promoter sequence-based order dependence.

4.4.2 Neutral DNA sequences do not reliably insulate promoters from upstream local DNA sequence context

In order to determine the shortest length of DNA spacer sufficient for insulating promoters from fluctuating upstream DNA sequence, two each of 12nt, 24nt and 36nt randomized DNA spacers were inserted upstream of constitutive promoters J23100, J23101 and J23104 as well as repressible promoter pTet (see Figure 4.10). Tested DNA spacer sequences were screened for promoter activity or for stable secondary structures that could potentially occlude access to the promoter and candidate spacer sequences with either were eliminated. The base composition of each pair of 12nt, 24nt and 36nt DNA spacer candidates were similar in their CG%. Insulated promoters were cloned into RFP expression cassettes with a medium-strong bicistronic design (BCD) RBS element BCD12 using MoClo DNA assembly. All expression cassettes were designed with MoClo fusion site “E” (5'-GCTT) upstream of the promoter. Circuits were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of red fluorescence measured. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

We found that for all promoters tested, the mean RFP expression of the two tested 12nt and 24nt spacers values did not fall within two standard deviations of each other (see Figure 4.10). However, the mean RFP expression values for the two tested 36nt spacers upstream of constitutive promoter J23100 (J23100-36INS-1 mean: 2.71×10^3 MEFLS; range: $2.65 \times 10^3 - 2.77 \times 10^3$ MEFLS, and 36INS-2 mean: 2.75×10^3 MEFLS, range: $3.00 \times 10^3 - 2.52 \times 10^3$) and repressible promoter pTet (pTet-36INS-1 mean: 1.12×10^3 MEFLS; range: $2.84 \times 10^3 - 444$ MEFLS, and pTet-36INS-2 mean: $1.08 \times 10^3 - 886$.) fell within two standard deviations of each

other and the two spacers could be considered equivalent. However, for constitutive promoter J23101 and J23104 the mean RFP expression for the two spacers tested were not within two standard deviations of each other.

4.4.3 Rationally designed “neutral” 36nt DNA spacer sequences affect promoter expression

The effect of the tested 36nt DNA spacer sequences when used to insulate against the 5' MoClo fusion site was tested (see Figure 4.12). Insulated versions of J23100_A and J23100_E RFP expression cassettes was assembled using MoClo and compared to each other and to the uninsulated J23100_A or J23100_E, respectively. To test, circuits were first grown in LB broth with appropriate antibiotics as described in Chapter 3, Section 3.3. Circuits were then diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of red fluorescence measured. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

Insulating J23100_A and J23100_E with the two tested rationally designed 36nt DNA spacers reduced the expression difference between upstream MoClo fusion site “A” (5'-GGAG) and “E” (5'-GCTT) from 3.5x to 1.34x (INS-1) and 1.19x (INS-2) respectively. The mean RFP expression between J23100_A and J23100_E for both tested DNA spacers was found to be within 2 standard deviations of each other and therefore not statistically significant. For each rationally designed DNA spacer tested, however, insertion of the spacer sequence upstream of promoter J23100 depressed RFP expression by 10.24x for J23100_A insulated with INS-1 and 9.93x for J23100_E insulated with INS-2 compared to the uninsulated version of the promoter. Mean RFP expression for J23100_E insulated with INS-1 and INS-2 were lower than

the uninsulated J23100_E by 3.91x and 3.37x, respectively.

4.4.4 Expanding the rationally designed J23100_A 36nt DNA spacer library did not yield DNA spacers providing consistent insulation

It was possible that the lower expression of the 36nt DNA spacers tested above compared to the uninsulated J23100 promoter was due to random chance. To check if this was true, the 36nt DNA spacer library was expanded from 2 DNA spacers to 10 spacers. The selection criteria for spacer sequences was kept identical to that described above in Section 4.4.2. Spacers were inserted 5' of promoter J23100 separating the promoter from its upstream MoClo fusion site. RFP expression cassettes were created using the expanded insulated J23100_A library using BCD12 as above using MoClo DNA assembly. To test, circuits were first grown in LB broth with appropriate antibiotics as described in Chapter 3, Section 3.3. Circuits were then diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of red fluorescence measured. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

Mean fluorescence expression values were compared to RFP expression of the uninsulated J23100_A promoter. With the exception of one DNA spacer (INS-10), which failed to clone, RFP expression cassettes were generated for all other DNA spacers (spacer sequences available in Table 3.8). All insulated J23100_A constructs showed depressed RFP expression compared to uninsulated J23100_A. Furthermore, INS4, INS-6, INS-8 and INS-9 cassettes demonstrated lower RFP expression compared to other insulated J23100_A cassettes while INS-5 and INS-7 showed high RFP expression compared to J23100_A cassettes insulated with INS-1, INS-2 and INS-3.

4.4.5 Rationally designed DNA spacers provide unreliable insulation

The inverters that were closest to each other in performance were selected to insulate promoters in the 24 permutations set of the test inverter. I selected 4 DNA spacers as the two instances of pBAD promoter in each inverter were to be insulated with separate yet equivalent DNA spacers to minimize artifacts from homologous recombination. Although pTet was found to be robust, we chose to insulate it for consistency throughout the device. The DNA spacers were selected from the tested insulators in Figure 4.12. The sequences for the insulated promoters can be found in Table 3.8.

Table 4.1: **Composition characteristics for rationally designed DNA spacer sequences.** Selected sequences were all 36 bp, had a CG% content of between 40% and 60%, formed no stable secondary structures and did not contain any known promoter activity.

Name	Sequence	TmC	CG%	nt	Promoter
INS-1	cctgcagtattcattttcagc ttacggaaggtagat	74.2	41.7	36	J23100
INS-2	cgttgttcgcgaccctgttc tgaggcaatggtgggc	86.9	58.3	36	pBAD
INS-3	tagcacgcttttctgcgtatg gttggtcagcttcta	79.1	47.2	36	pBAD
INS-5	cggagagaatacatagtaaa gaaacctcactgttgt	70.8	38.9	36	pTet

Using these insulated promoter parts, we rebuilt and tested our 24 inverter permutation set using MoClo assembly. The inverters were grown overnight in LB Broth as described in Chapter 3 and then grown for a second night in the presence (ON state) or absence (OFF state) or 1 mM L-arabinose. The inverters were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green and red fluorescence was measured by flow cytometry. All sam-

ples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

We found that while insulating the promoters in our 24 inverter set did eliminate the pattern of high GFP expression in inverter devices where the pBAD-GFP TU was flanked by the assembly scar “A” (pBAD-GFP_A), it did not regularize the expression profile of the devices or relieve the variations between inverters.

4.5 Conclusions

I show here that genetic inverters equivalent in design behave unpredictably resulting in faulty device behavior (See Figure 4.2). I have also shown that this performance unpredictability is dependent on the order of transcription units within the device. In a comparison of the GFP and RFP fold change profiles of 24 equivalent inverters, I found that inverter devices in which the pBAD-GFP was the first transcription units transcription unit showed noticeably higher GFP expression compared to when pBAD-GFP was the second, third or fourth transcription unit in the device. Promoter parts appear to be sensitive to changes in DNA sequence at their 5' boundaries.

Most commonly used BioBrick [35] promoter parts were originally designed without any buffering sequences between key protein binding sequences like the -35 RNA Polymerase (RNAP) binding sites in minimal constitutive promoters or the transcription factor binding sites for regulated promoters and have been used as such in the creation of engineered genetic devices. When these promoter parts are incorporated into new combinations and into new plasmid vector backbones, the sequence at the 5' junctions of the promoters is altered. In the absence of buffering sequences at the 5' promoter boundaries, these altered 5'promoter junctions could impact protein binding and thus result in a different level of expression than in the original context in

which the promoter was characterized. An investigation into the mechanism of such variations in promoter activity as a result of altered genetic context is beyond the scope of this study, but would certainly provide valuable information for improved synthetic promoter part design for new synthetic biology parts.

Neutral DNA sequences (spacers or insulators) have previously been used to buffer promoter parts from upstream DNA sequences with transcription activating properties in bacteria [16]. However, these upstream transcription activating sequences are traditionally -60 to -40 base pairs upstream of the transcription start site (TSS). DNA spacers have not been used previously as a tool to insulate against sequence variations immediately upstream a defined promoter part. As such, there were no available parameters for successful spacer design. I defined a “neutral” DNA sequence as one that matched the base composition of intergenic regions in *E. coli* between operons [48]. Through my preliminary work on DNA spacer selection, I concluded that 36 bp DNA spacers were often sufficient for insulating promoters from expression variations caused by changes in the DNA sequence at the 5' promoter boundary.

My preliminary work on the selection of DNA spacers made it clear that designing, testing and selecting DNA spacers would prove a laborious process. Although a very small number of DNA spacers were tested, the DNA base composition for the spacers with the most similar expression profiles did not fall within a narrow range by which to establish any selection criteria. Without a more nuanced understanding of sequences that would provide reliable and adequate insulation, designing and testing spacers individually appeared to be a laborious and fruitless process. At this point, I abandoned this approach for a newer methodology that is described in Chapter 3, Section 3.2.3.1 and shown in Figure 3.3.

Figure 4.2: (Caption continued on following page.)

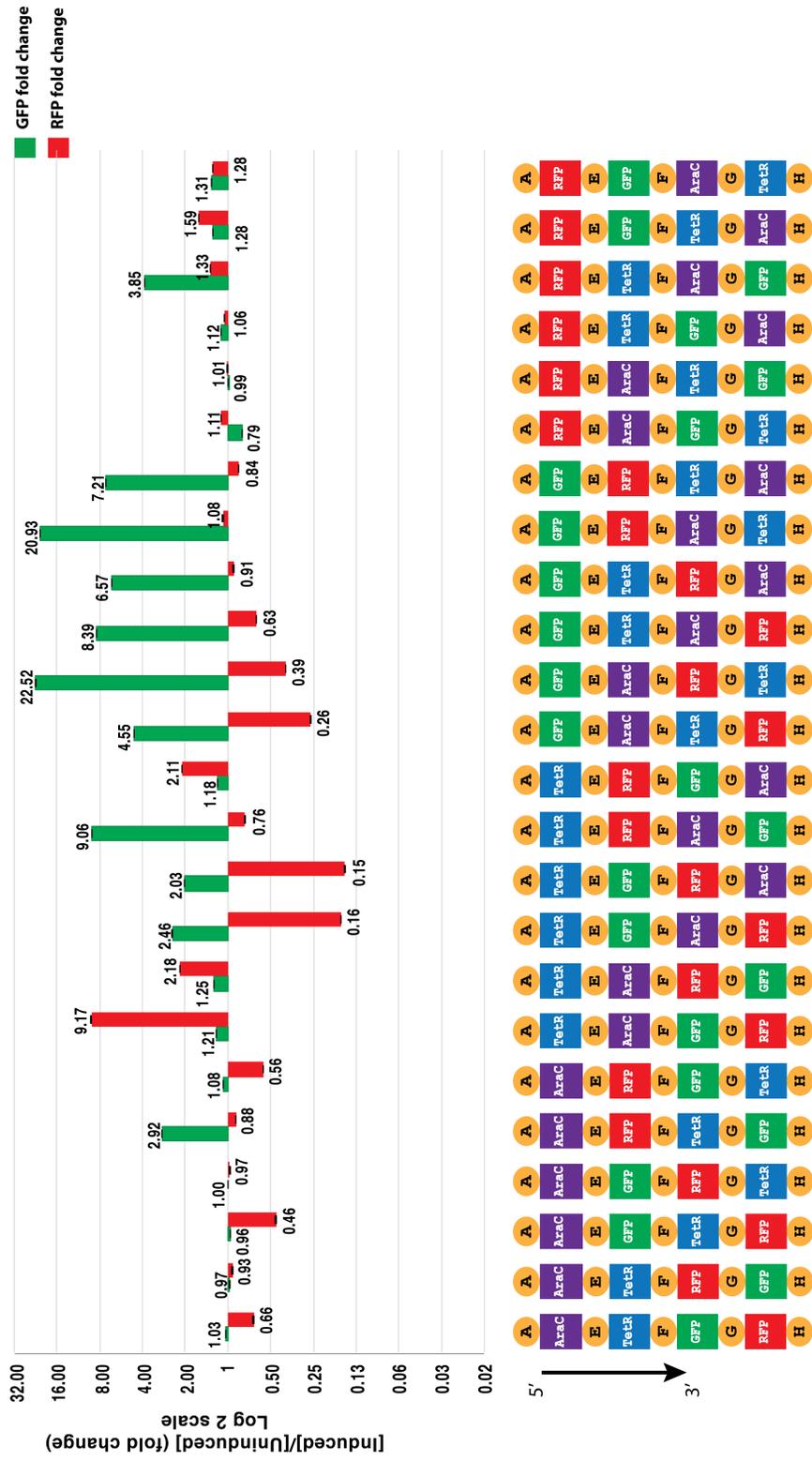


Figure 4.2: **Expression profile of the 24 uninsulated inverter permutations of our test TetR inverter upon induction with L-arabinose.** (Continued from previous page.) Assembly scars upstream of promoters cause anomalous expression and order dependence of modules in genetic devices assembled using Golden Gate-derived DNA assembly methods. Comparison of GFP and RFP fluorescence fold changes upon induction with L-arabinose of the 24 inverter permutation set in which no promoters are insulated reveals device-to-device variations and a clear pattern of high GFP fluorescence in inverters where pBAD-GFP_A was TU1 in the inverter configuration. All fluorescence readouts were converted to absolute units (MEFLs). Conversion of fluorescence readouts to MEFLs was done using the TASBE tools by our collaborators.

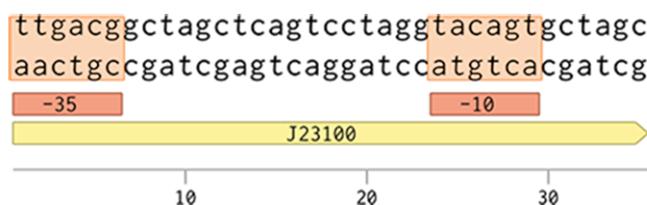


Figure 4.3: **J23100 promoter architecture.** Promoter J23100 is a 35 base pair minimal constitutive promoter consisting of -35 and -10 RNA polymerase binding sites separated by 17 base pairs. The J23100 promoter does not contain any DNA sequences that buffer the -35 RNA polymerase sigma binding site from fluctuating DNA sequences upstream of its 5' boundary. Promoter J23100 sequence is shown in yellow. -35 and -10 RNAP binding site sequences are shown in orange.

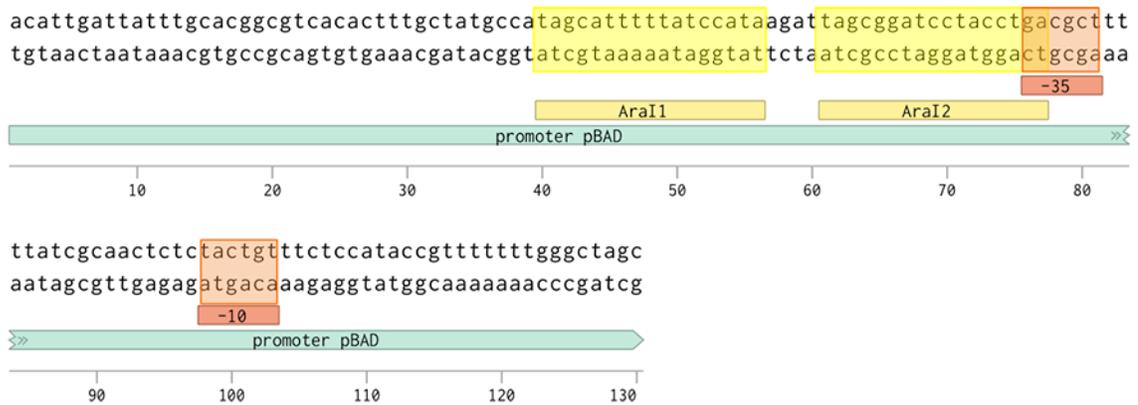


Figure 4.4: **pBAD promoter architecture.** Promoter pBAD is a 130 base pair inducible promoter consisting of two operator sequences that bind the repressor, AraC, in addition to the -35 and -10 RNA polymerase sigma subunit binding sites. The pBAD promoter contains 39 base pairs between the 5 boundary of the promoter and the first protein binding site. Promoter pBAD sequence is shown in light blue. -35 and -10 RNAP binding site sequences are shown in orange, and the AraC operator sites are shown in yellow.

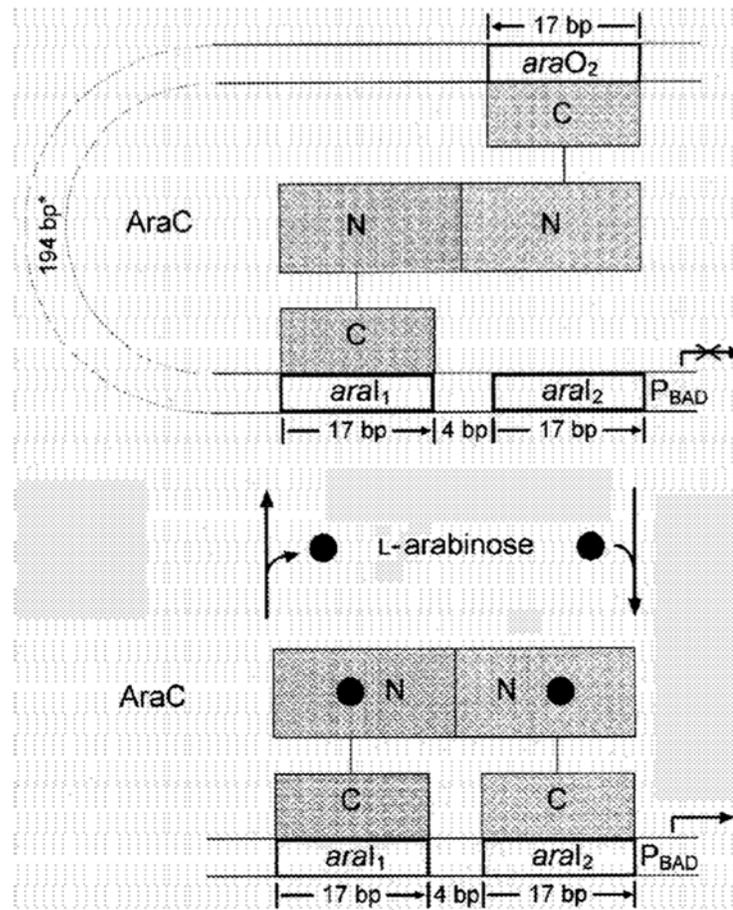


Figure 4.5: **Model of AraC induction by L-arabinose.** A dimer of AraC binds to the operators I1 and O2 in the repressing state (N and C denote the N and C-terminal domains of AraC, respectively). Upon addition of the inducer L-arabinose (represented by filled circles), AraC undergoes a conformational change that makes long-distance looping impossible and so restricts binding to the proximal target sites I1 and I2 . The disappearance of the DNA loop and the occupation of the I2 site lead to activation at P_{BAD} . The depicted sizes and distances are not drawn to scale. An asterisk (*) indicates the intervening base-pairs between araI1 and araO2. Figure and figure legend reproduced from How AraC Interacts Specifically with its Target DNAs, by P. Niland *et al.*, Journal of Molecular Biology, 1996. [52]

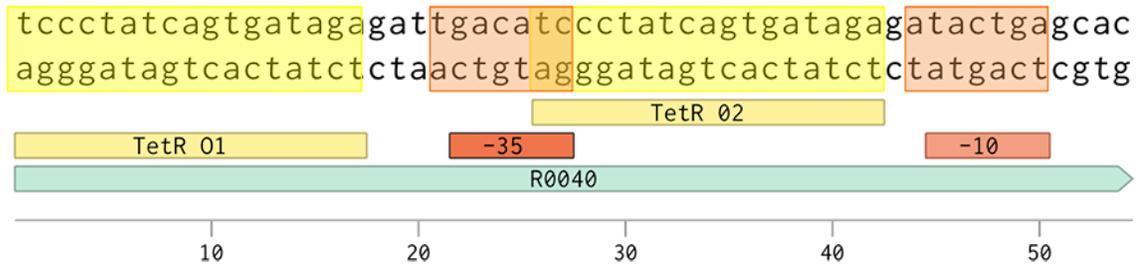


Figure 4.6: **pTet promoter architecture.** Promoter pTet is a 54 base pair repressible promoter consisting of two operator sequences (TetR-O₁ and TetR-O₂) that bind the repressor, TetR, in addition to the -35 and -10 RNA polymerase sigma subunit binding sites. The pTet promoter does not contain any DNA sequences that buffer the TetR-O₁ from fluctuating DNA sequences upstream of its 5' boundary. Promoter pTet sequence is shown in light blue. TetR-O₁ and TetR-O₂ sites are shown in yellow and the -35 and -10 RNAP binding site sequences are shown in orange.

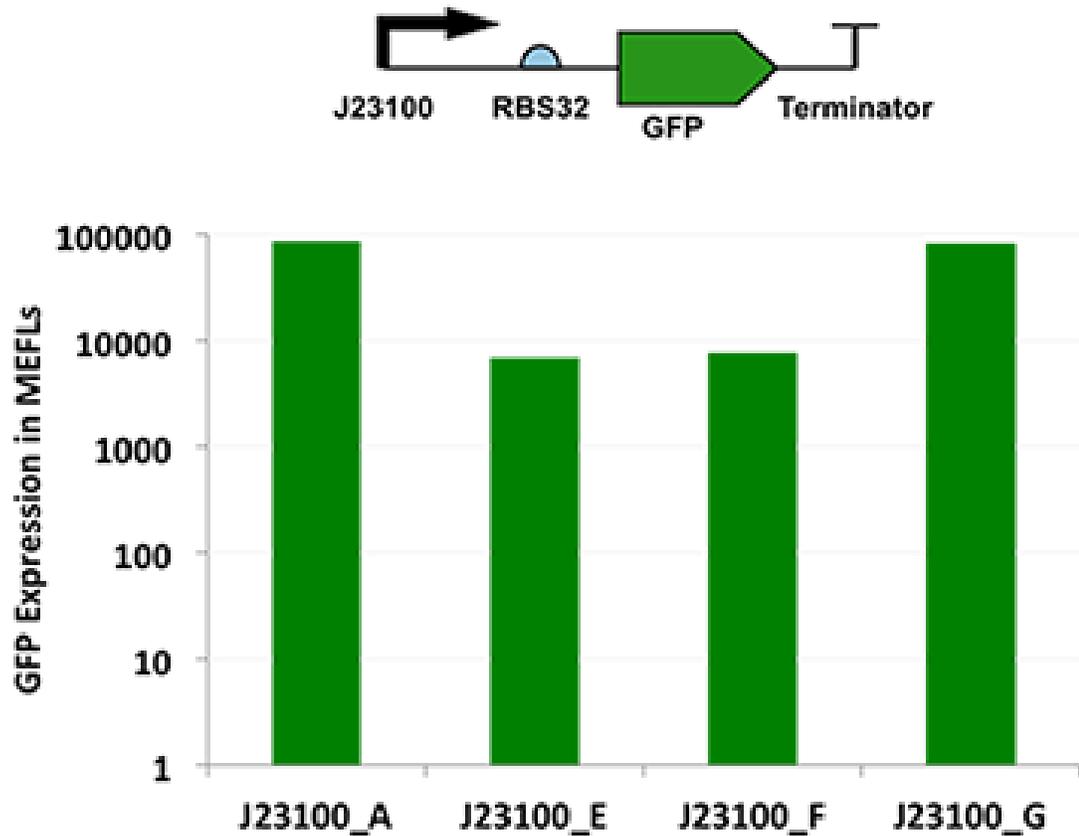


Figure 4.7: **Constitutive promoter J23100 expression affected by upstream DNA sequence.** Individual J23100-GFP expression cassettes were built using J23100_A, J23100_E, J23100_F and J23100_G using the RBS B0032m_BC and double terminator B0015 with the appropriate 3' fusion sites as listed in Table 3.5. GFP fluorescence was used to measure promoter expression. J23100_A and J23100_G had similar expression levels, while both J23100_E and J23100_F expression were >10x lower than J23100_A expression. Altering 4 bp 5' of promoter J23100 GGAG (J23100_A) to GCTT (J23100_E) or CGCT (J23100_F) lowered expression GFP expression from promoter by 12.5x and 11x respectively. GFP expression was not significantly altered when upstream fusion site was modified to TGCC (J23100_G). All fluorescence measurements are in MEFLs.

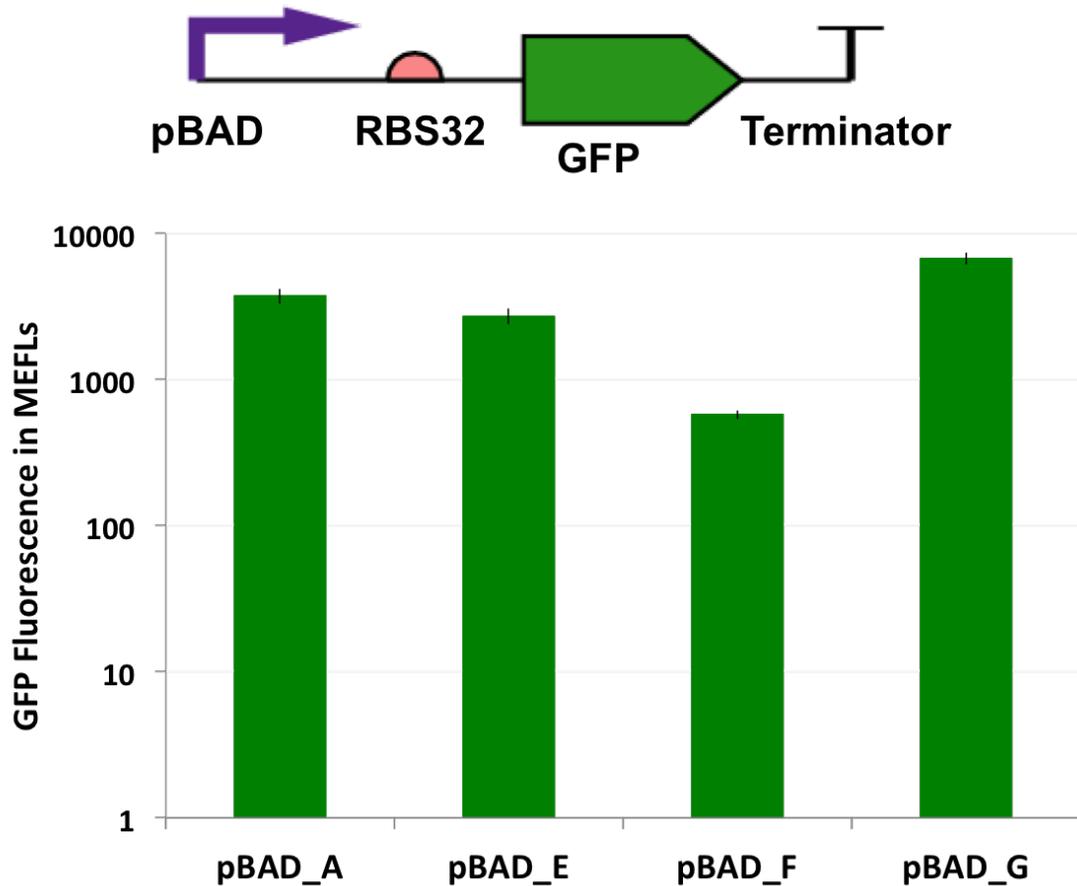


Figure 4.8: **Inducible promoter pBAD expression affected by upstream DNA sequence.** Individual pBAD-GFP expression cassettes were built using pBAD_A, pBAD_E, pBAD_F and pBAD_G using the RBS B0034m_BC and double terminator B0015 with the appropriate 3' fusion sites as listed in Table 3.5. GFP fluorescence was used to measure promoter expression. pBAD_A, pBAD_E and pBAD_G expression levels were within 2-folds of each other. pBAD_F expression was 6.50x lower than pBAD_A expression. Altering 4 bp 5' of promoter pBAD from GGAG (pBAD_A) to CGCT (pBAD_G) increased GFP expression from promoter by 1.44x. All fluorescence measurements are in MEFLs.

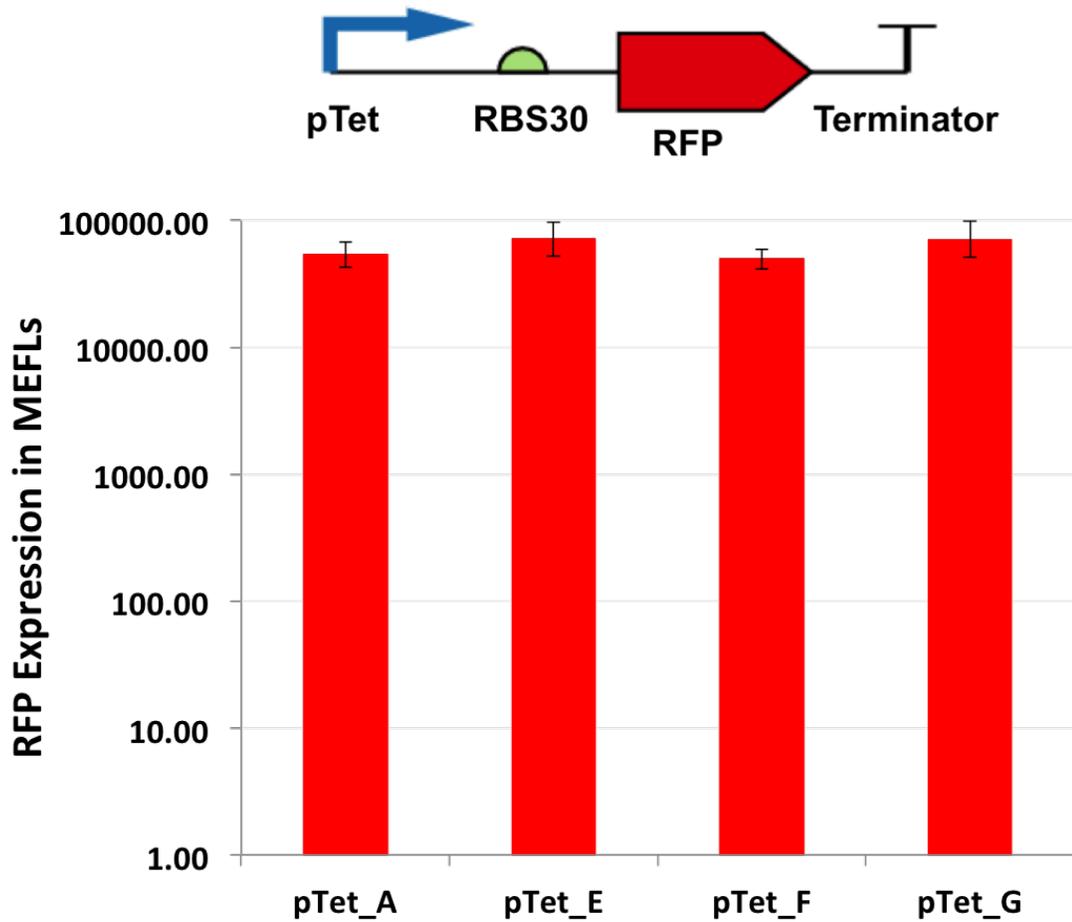


Figure 4.9: **Repressible promoter pTet expression is unaffected by changes in its upstream DNA sequence** Individual pTet-RFP expression cassettes were built using pTet_A, pTet_E, pTet_F and pTet_G using the RBS B0030m_BC, E1010_CD and double terminator B0015 with the appropriate 3' fusion sites as listed in Table 3.5. RFP fluorescence was used to measure promoter expression. The expression from the four MoClo fusion site variants of pTet-RFP were statistically indistinguishable from each other. All fluorescence measurements are in MEFLs.

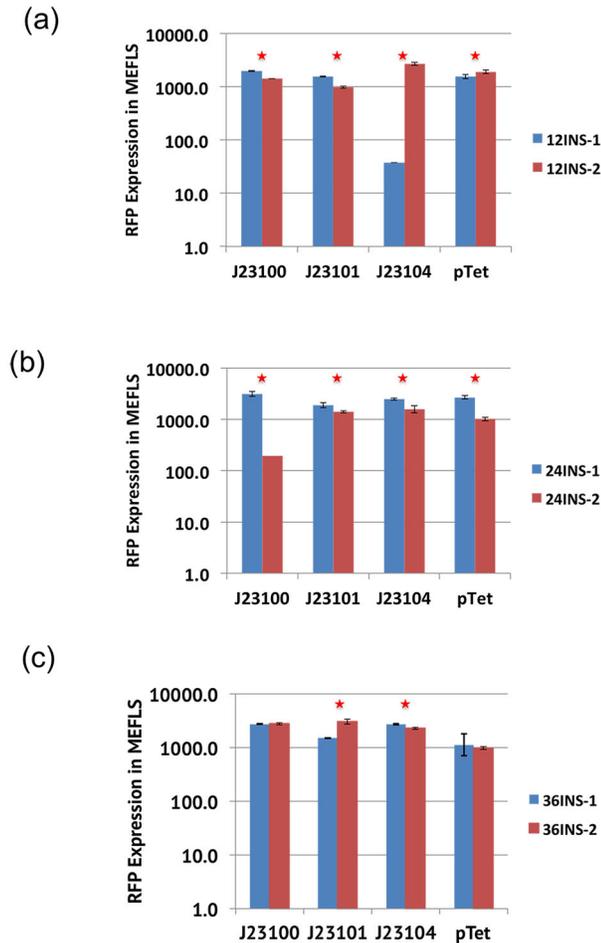


Figure 4.10: **RFP expression from promoter cassettes insulated with 12nt, 24nt and 36nt DNA spacer sequences** Three DNA spacer lengths were tested to determine minimum spacing required for promoter insulation. Two spacers of each length were tested to reduce the chances of homologous recombination if multiple insulated promoters need to be incorporated into a single target device. Equivalence of DNA spacers was measured by comparing mean RFP expression of the two insulated versions of each promoter. All measurements are in MEFLs. (a) The two sampled 12nt spacers do not reliably insulate any tested promoter. (b) The two sampled 24nt spacers do not reliably insulate any tested promoter. (c) The two tested 36nt spacers reliably insulate constitutive promoter J23100 and repressible promoter pTet but not constitutive promoters J23101 or J23104.

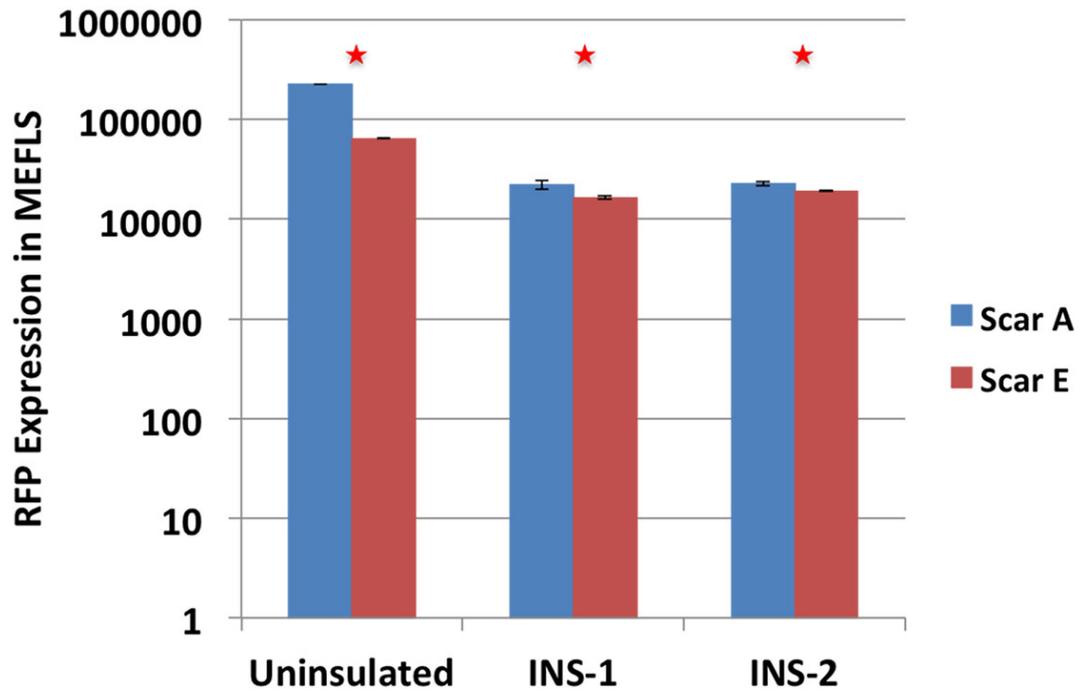


Figure 4.11: **Insulated promoters reduce dependence on upstream DNA sequence context at the cost of altered promoter expression levels** J23100_A or J23100_E were insulated with the two previously tested rationally designed 36nt DNA spacers. Insulated promoters were assembled into RFP expression cassettes with bicistronic design (BCD) RBS element BCD12. Insulated J23100 promoters show 10x lower expression than the uninsulated promoter. Insulating promoters with 36nt DNA spacers reduce the upstream DNA sequence-dependent expression changes but do not eliminate them entirely. Insulated promoter J2310. All measurements are in MEFLs.

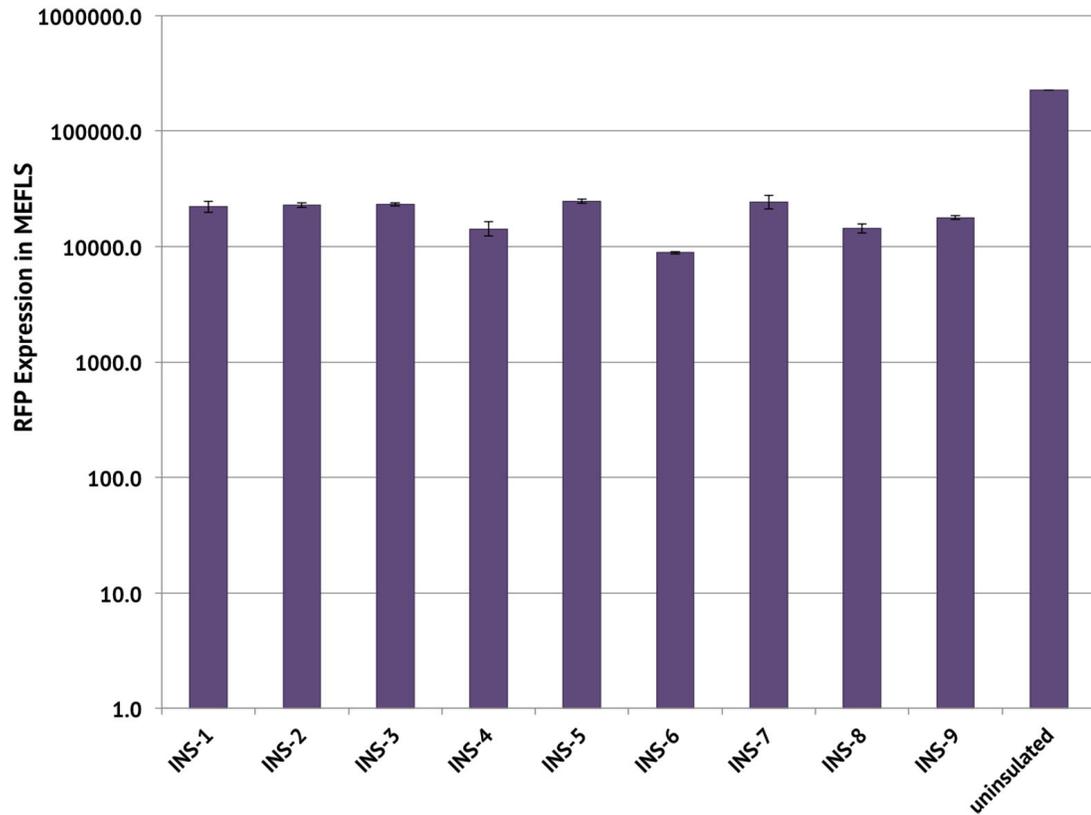


Figure 4.12: **Expanding the rationally designed 36nt DNA spacer library showed variability in insulated J23100_A expression** J23100_A was insulated with DNA spacers INS1-INS10. Insulated promoters were assembled into RFP expression cassettes with bicistronic design (BCD) RBS element BCD12. The expanded insulated J23100 promoter library continues to show approximately 10x lower expression than the uninsulated J23100_A promoter. Insulated promoter J2310 All measurements are in MEFLs.

Figure 4.13: (Caption continued on following page.)

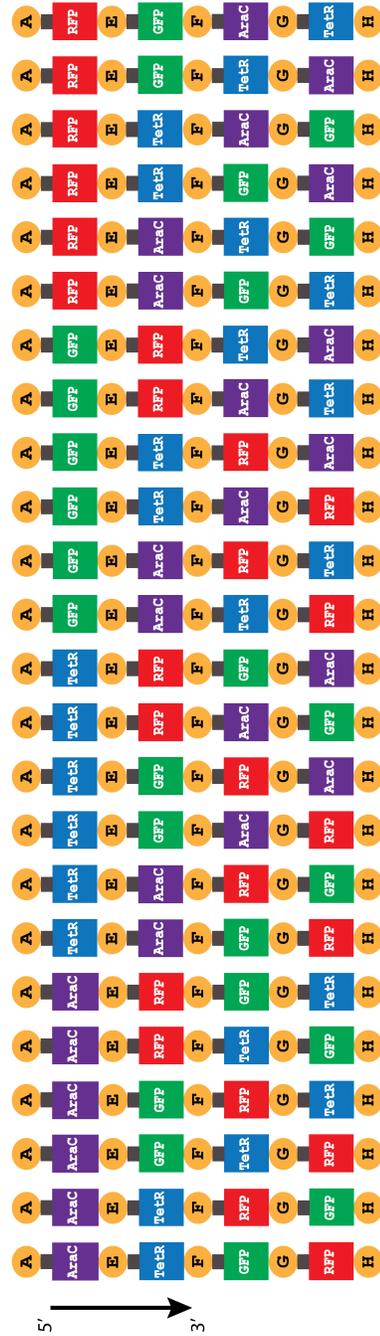
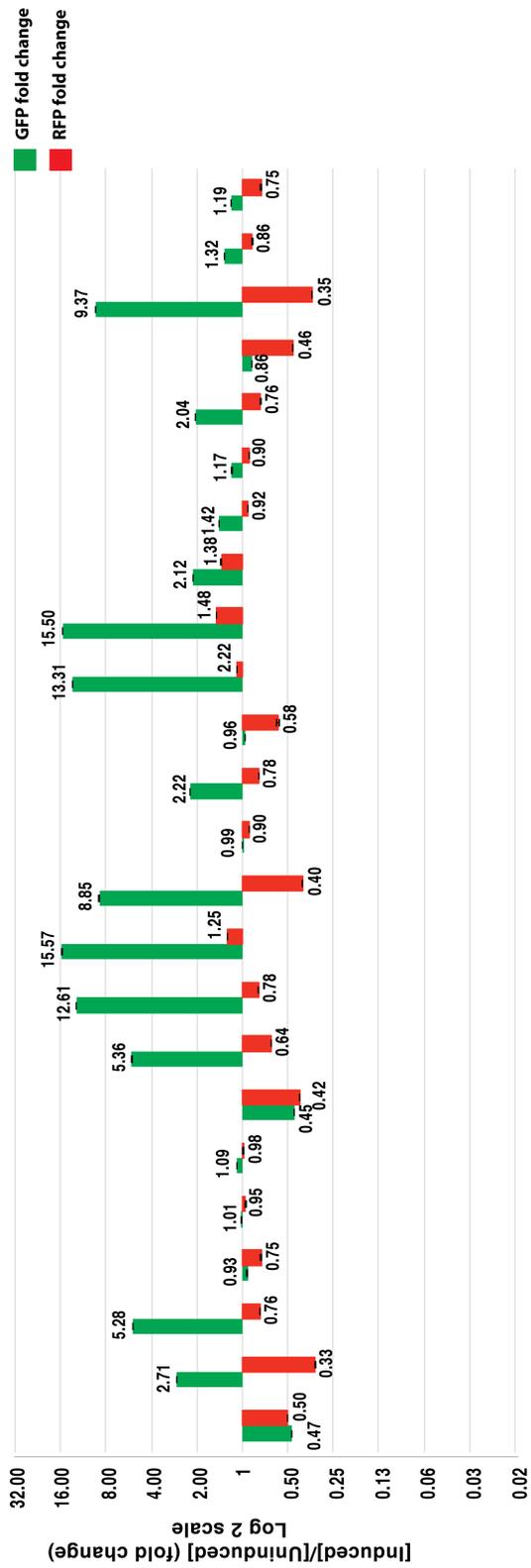


Figure 4.13: **Expression profile of 24 inverter set with all promoters insulated with rationally selected DNA spacers upon induction with 1mM L-arabinose.** (Continued from previous page.) Rationally designed DNA spacers do not guarantee reliable insulation. Comparison of GFP and RFP fluorescence fold changes upon L-arabinose induction of 24 inverter permutations built with rationally designed insulated promoters removes previously observed high GFP fold change pattern from the uninsulated inverter set, but does not eliminate device to device expression fluctuations. All fluorescence readouts were converted to absolute units (MEFLs).

Chapter 5

DNA Spacers: High-throughput screening of insulated promoter libraries

5.1 Introduction

In this chapter I present my results from my novel DNA spacer screening method. I screened almost 2000 samples from insulated J23100_A, J23100_E, J23100_H and J23100_K libraries (a fusion site that is not part of my inverter designs) to compare patterns of distribution of insulated expression cassettes. I further apply my spacer screening methodology to the other promoters in my test inverter, pBAD and pTet. As promoters pBAD and pTet are regulatable and have transcription factor binding sites that a good DNA spacer candidate should not disrupt, I chose to evaluate them in a way such that I could verify that the DNA spacer I selected from my library did not have any impact on transcription factor binding. In fact, the insulated library for promoter pTet was generated within the context of one of the permutations of the test inverter. The detailed methodology is described in Section 3.2.3.1 and a schematic representation is shown in Figure 3.3. Briefly, a library of insulated promoter expression cassettes upstream of any chosen promoter could be generated by performing an inverse PCR reaction using 100% degenerate primers followed by a blunt-end ligation to circularize the linear PCR products. The library is then

transformed into *E. coli* (Section 3.2.5 and screened by flow cytometry against a reference device to obtain DNA spacers that provide the precise level of insulation needed.

5.2 Results

5.2.1 Distribution of 36nt-insulated J23100 expression cassettes

Using our new DNA spacer library generation and screening method described in Section 3.2.3.1, we created and screened approximately 2000 total samples from 36nt-J23100_A, 36nt-J23100_E and 36nt-J23100_H (fusion site not in 5' of 24 inverter set designs; 5'- ACTA) and 36nt-J23100_K (fusion site not in 5' of 24 inverter set designs; 5'- AGTC) sequences to sample the range of expression level distribution for 36nt DNA spacer sequences placed upstream of promoter J23100, and also to screen for a DNA spacer to use upstream of promoter J23100 in the 24 inverter set designs.

Insulated J23100 expression cassette libraries were generated as described in Chapter 3, Section 3.2.3.1 and shown in Figure 3.3. Each colony from a library transformation plate represents a distinct insulated promoter cassette. Circuits were first grown in LB broth with appropriate antibiotics as described in Chapter 3, Section 3.3. Circuits were then diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green fluorescence measured. As this was a screen, samples were not in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9]. My collaborators assisted me with plotting the expression distribution for the insulated promoter libraries.

We observed a very wide range of expression levels as can be seen in Figures 5.1, 5.2, 5.3 and 5.4, respectively. Addition of some spacers greatly reduced GFP

expression while others increased it.

5.2.1.1 Distribution of GFP expression of screened samples from 36nt-J23100_A-GFP DNA spacer library

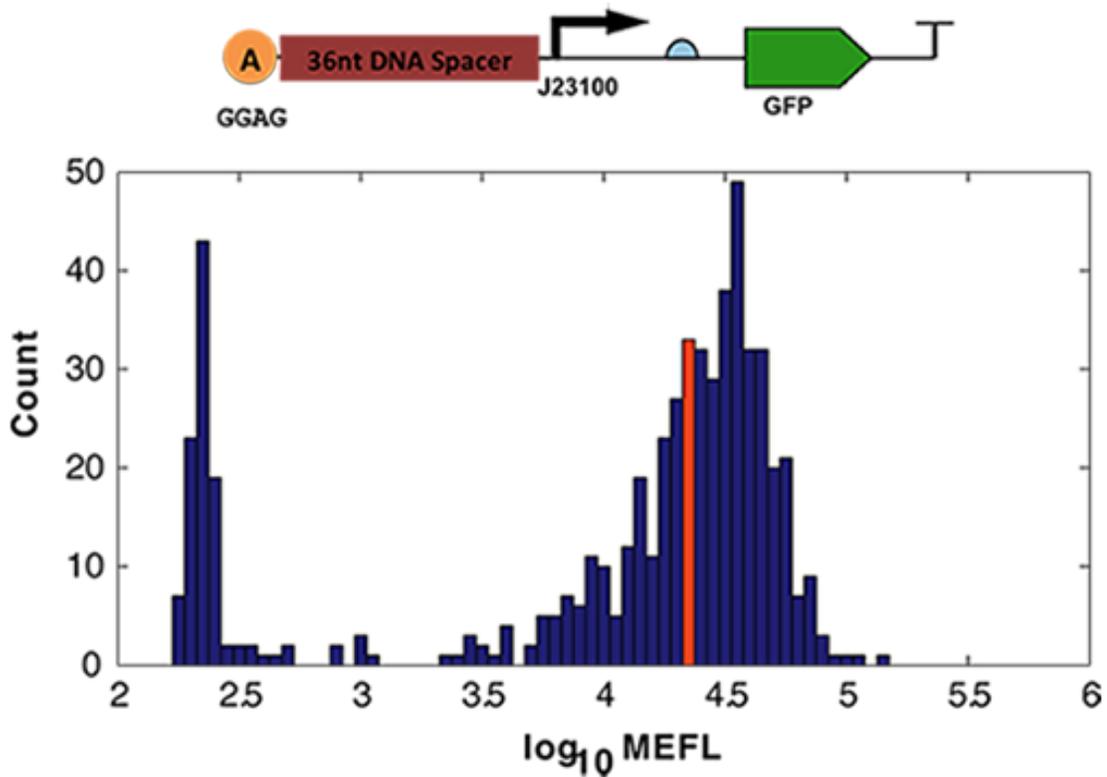


Figure 5.1: **Distribution of insulated 36nt-J23100_A GFP library.** Expression of approximately 1200 samples from the 36nt-J23100_A spacer library were checked by flow cytometry for GFP expression levels. The orange bar marks the expression level of insulated J23100_A cassettes with the same expression level as the uninsulated J23100_A-GFP reference circuit. The expression distribution contained a clear peak. All fluorescence readouts were converted to absolute units (MEFLs).

Figure 5.1 shows the distribution of approximately 1200 36nt insulated J23100_A-GFP cassettes. The expression of the reference control (uninsulated J23100_A-GFP; 2.45×10^4 MEFLs) fell within the body of the bulk of the distribution but was lower

than peak insulated cassette expression. Most spacers were found to exert a strong genetic context effect on the J23100 promoter. In approximately 100 of the tested insulated J23100_A samples, the addition of a spacer abolished expression to almost negative control autofluorescence levels.

5.2.1.2 Distribution of GFP expression of screened samples from 36nt-J23100_E-GFP DNA spacer library

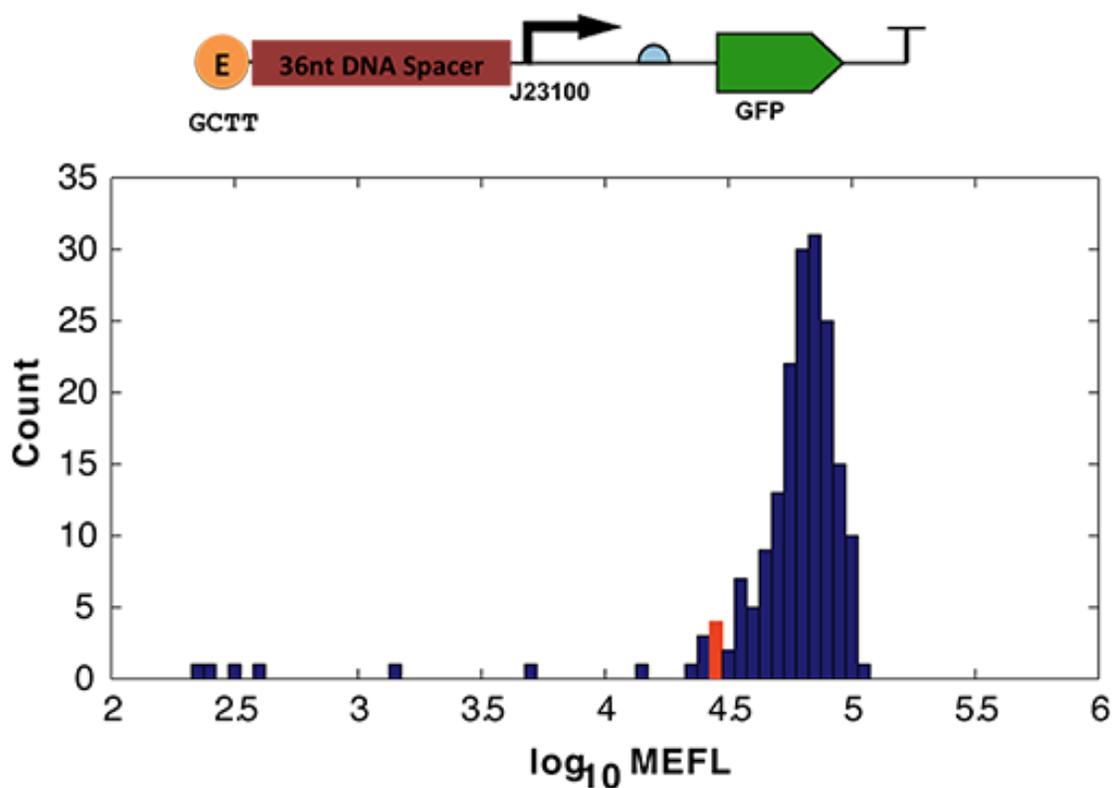


Figure 5.2: **Distribution of insulated 36nt-J23100_E GFP library.** Expression of approximately 400 samples from the 36nt-J23100_E spacer library were checked by flow cytometry for GFP expression levels. The orange bar marks the expression level of insulated J23100_E cassettes with the same expression level as the uninsulated J23100_A-GFP reference circuit. The expression distribution had a clear peak. All fluorescence readouts were converted to absolute units (MEFLs).

Figures 5.2, 5.3 and 5.4 show the distribution of approximately 400 36nt insulated J23100_X-GFP cassettes in 3 cases. The expression of the reference control (uninsulated J23100_A-GFP; 2.45×10^4 MEFLs) fell within the body of the bulk of the distribution but was lower than peak insulated cassette expression. Most spacers were found to exert a strong genetic context effect on the J23100 promoter. In approximately 10 of the tested insulated J23100_E samples, the addition of a spacer abolished expression to almost negative control autofluorescence levels.

5.2.1.3 Distribution of GFP expression of screened samples from 36nt-J23100_H-GFP DNA spacer library

Figure 5.3 shows the distribution of approximately 200 36nt insulated J23100_H-GFP cassettes. While the expression distribution contained a clear peak, the GFP expression level of the reference control (uninsulated J23100_A-GFP; 2.45×10^4 MEFLs) fell within the body of the bulk of the distribution but was lower than peak insulated cassette expression. Most spacers were found to exert a strong genetic context effect on the J23100 promoter. In approximately 10 of the tested insulated J23100_H samples, the addition of a spacer abolished expression to almost negative control autofluorescence levels.

5.2.1.4 Distribution of GFP expression of screened samples from 36nt-J23100_K-GFP DNA spacer library

Figure 5.4 shows the distribution of approximately 200 36nt insulated J23100_K-GFP cassettes. The expression of the reference control (uninsulated J23100_A-GFP; 2.45×10^4 MEFLs) fell at the lower end of the distribution of expressions and was lower than peak insulated cassette expression. Most spacers were found to exert a strong

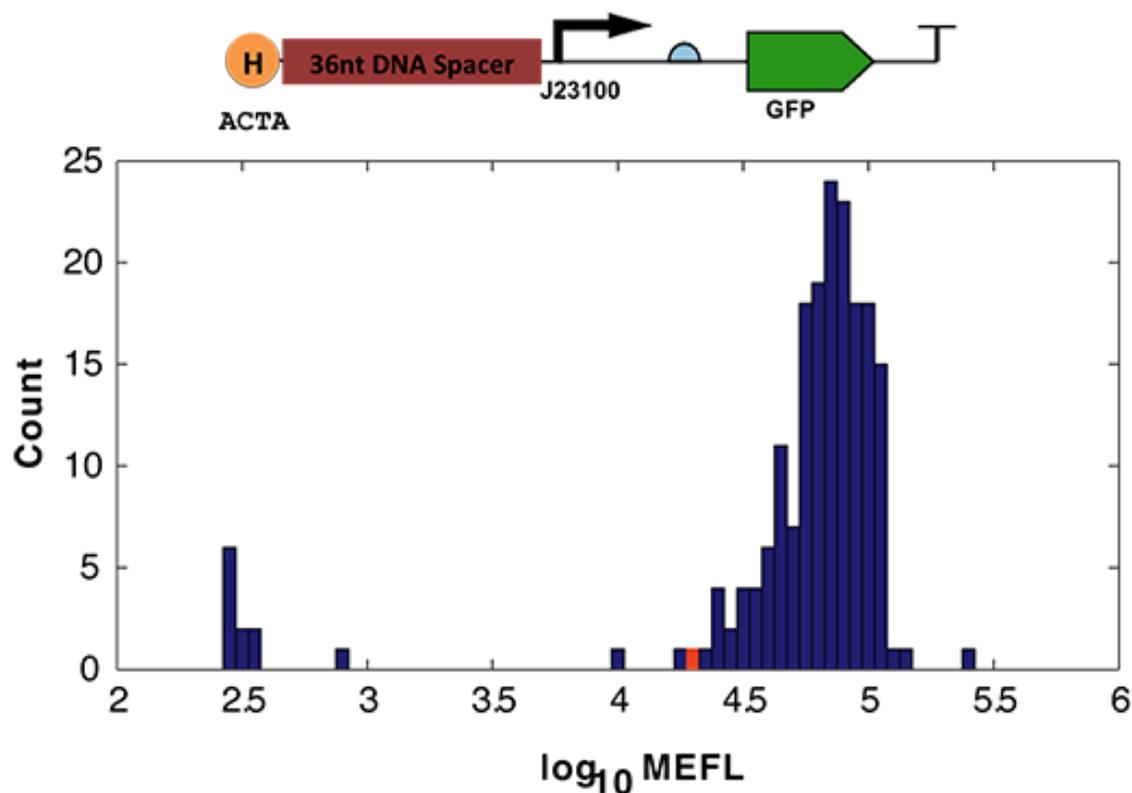


Figure 5.3: **Distribution of insulated 36nt-J23100_H GFP library.** Expression of approximately 400 samples from the 36nt-J23100_H spacer library were checked by flow cytometry for GFP expression levels. The orange bar marks the expression level of insulated J23100_H cassettes with the same expression level as the uninsulated J23100_A-GFP reference circuit. The expression distribution contained a clear peak. All fluorescence readouts were converted to absolute units (MEFLs).

genetic context effect on the J23100 promoter. In approximately 10 of the tested insulated J23100_K samples, the addition of a spacer abolished expression to almost negative control autofluorescence levels.

5.2.2 Analysis of DNA spacer sequences from J23100_X screens

Approximately 600 insulated cassettes were sequenced in an attempt to investigate the characteristics of the spacers in the distribution. Of the 600 spacers sequenced,

we found 2 duplicates from 36nt-J23100_A and 36nt-J23100_E plates (see Figures 5.5 and 5.6). The spacer libraries from which the samples were screened were created and tested on separate days, so the result is unlikely to be due to contamination. The CG% of all sequenced spacers ranged from 27.8% to 66.7%. All screened spacer sequences that were not exactly 36nt in length, or in which promoter or upstream MoClo fusion site was either missing or mutated were eliminated. Our collaborators assisted me in presenting this result for visualization.

After trimming, 150 spacer sequences remained from the 600 sequenced samples. The CG% was recalculated for the trimmed set, but both the mean and median CG% remained unchanged. The modal expression level of the distribution remained unchanged. Furthermore, the mean and median CG% of sequenced spacers were 42.7% and 41.7%, respectively. The modal CG% for these was also 41.7%. A search for consensus bases or motifs within the trimmed set using the WebLogo 2.8.2 online tool [65] revealed no consensus patterns. However, we did find that among the trimmed set of DNA spacers, the final base in the 36nt spacer sequences was predominantly (deoxy)thymine (T). As such, the screened spacers did not differ significantly from the rationally designed DNA spacer sequences in Table 4.1.

5.2.3 Distribution of PCR-generated 36nt-insulated pBAD library

The insulated pBAD expression cassette library was generated as described in Chapter 3, Section 3.2.3.1 and shown in Figure 3.3. Each colony from a library transformation plate represents a distinct insulated promoter cassette. Circuits were first grown in LB broth with appropriate antibiotics as described in Chapter 3, Section 3.3. Circuits were then diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green fluorescence measured. As

this was a screen, samples were not in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9]. My collaborators assisted me with plotting the expression distribution for the insulated promoter libraries.

Using the new PCR-based DNA spacer library generation method utilizing degenerate primers described in Section 3.2.3.1, a library of 36nt-insulated pBAD_F promoter variants were generated in *E. coli*. MoClo fusion site “F” (5'-CGCT) was chosen as the upstream DNA sequence because it caused the largest reduction in pBAD expression out of the four tested MoClo fusion sites. In order to avoid DNA spacers that disrupted transcription factor (AraC) binding to the pBAD promoter, a reference and PCR template genetic circuit was selected that would allow measurement of pBAD-GFP expression before and after induction. Insulated pBAD samples that failed to show any GFP induction were discarded from consideration. Distribution of the insulated pBAD_F library was much tighter than the distribution of the insulated J23100 libraries for all tested upstream MoClo fusion sites. The GFP expression fold change of the reference circuit was 6.28x. Insulated pBAD_F cassettes showing a fold change greater than 10x are depicted by the orange bars in Figure 5.7. Two distinct but equivalent DNA spacers from the candidate spacers with GFP induction levels identical to that of the reference device were selected for insulating the two instances of the pBAD promoter in our 24 inverter permutation sets.

5.2.4 Distribution of PCR-generated 36nt-insulated pTet library

Unlike the promoter J23100 and pBAD, the repressible promoter pTet in isolation was found to be unaffected by variations in the DNA sequence at its 5' boundary.

As a template for spacer library generation, I selected an inverter configuration

where the pTet promoter was flanked by the MoClo fusion site “A” (5'-GGAG), and the J23100 and pBAD promoters were insulated. The inverters were grown overnight in LB Broth as described in Chapter 3 and then grown for a second night in the presence (ON state) or absence (OFF state) or 1 mM L-arabinose. The inverters were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, section 3.3.2 and the expression of green and red fluorescence was measured by flow cytometry. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

Figure 5.8 shows the distribution of RFP expression of insulated pTet promoters created within the context of a full inverter circuit. As RFP is inhibited when the inverter is induced with L-arabinose, a good candidate DNA spacer should not diminish RFP repression of the template inverter. RFP fold change on inverter induction of the template/reference circuit was 6.2x (red bars); the orange bars indicate the insulated pTet circuits with >10x repression of RFP expression upon induction. The distribution of RFP expression of the insulated pTet promoters was very wide and did not contain a clear single peak. Upon induction, in many insulated pTet promoters the RFP expression increased instead of decreasing upon induction. In several cases, RFP expression remained unchanged. However, in some devices, in over a tenth of the devices sampled, we observed that the inserted DNA spacer actually increased RFP repression upon induction of the inverter (see Figure 5.8(c)).

I received assistance from our collaborator in creating a visual representation of the insulated pTet promoter library.

5.3 Discussion

The wide distribution of insulated J23100 promoter expression shows that it may not be possible to rationally design truly neutral DNA spacers that provide precisely the level of insulation required without undergoing exhaustive trial and error. For minimal promoters like the BioBricks J231XX series [34] that are defined as the -35 and -10 RNA Polymerase (RNAP) binding sequences with an intervening 17 base pair sequence without any buffering sequences 5' of the -35 RNAP binding site, the use of insulated promoters may be critical to obtaining expression matching the promoter characterization data when recomposed into a larger synthetic genetic device.

5.3.1 Designing DNA spacers of alternate lengths

In my preliminary work on DNA spacers I show that 36 base pair spacers are sufficient for insulating promoter J23100. However, given the wide variation of expression levels of insulated DNA spacers observed during screening candidate spacers, it is evident that I was fortunate in my original selection of the two 36 base pair DNA spacers to test. It is quite possible that other lengths - both shorter and longer - can also provide reliable and precise promoter insulation.

Procedurally, there are no barriers to designing 100% degenerate DNA spacers of lengths shorter than 36 base pairs. Shorter DNA spacers can be obtained simply by reducing the number of degenerate nucleotides in each primer to one half of the final desired spacer length. However, there is an upper limit on the length of the spacer that can be inserted upstream of a promoter. The maximum length of a DNA spacer possible using this method is 50 nucleotides. This limitation is not due to the method itself; the current available DNA synthesis technologies do not allow companies to synthesize a stretch of 100% degenerate nucleotides longer than 25 nucleotides at the

moment. As technologies improve, longer DNA spacers will become possible. The cost of primers can also be reduced by using DNA spacers shorter than 36 base pairs.

5.3.2 Resource Cost of Insulating a Promoter

The most significant costs involved in insulating a promoter are the cost of primers, which are unlikely to be re-usable in a new context, as well as the time in lab that is required to generate the DNA spacer library and perform the screening. The biggest cost in insulating a promoter is the cost of the polyacrylamide gel electrophoresis (PAGE)-purification to each promoter in its final context. PAGE-purified primers are recommended for modified oligonucleotides (oligos) as well as oligos that are 60 nucleotides or longer. PAGE-purified primers cost \$110.0 per primer. However, it is possible to obtain a library of DNA spacers using cheaper HPLC-purified primers, bringing the cost of oligonucleotide primers down from \$110.0 per primer to \$40.0 per primer. Whether or not the distribution of expression of the spacer library obtained using HPLC-purified primers is different from that obtained using PAGE-purified primers has not yet been tested. Similarly, commercially prepared chemically competent cells (See Chapter 3) were used for transformation of the DNA spacer library. Transformations can be done more cheaply using homemade competent cells. The cost of insulating a single promoter as per the protocols in this thesis are provided in Table 5.1.

The process of generating an insulated promoter library and performing the screen for a DNA spacer adds 8-10 days to the build cycle before the testing phase, depending on the type and length of induction necessary, not including the time required to order a new promoter part with the DNA spacer selected by screening. The minimum time required to obtain a DNA spacer by the spacer library generation and

Table 5.1: **Approximate reagent cost to insulate a promoter.**

Reagent	Cost
Primers (PAGE-purified)	\$220
PCR reagents	\$4.24
PCR template elimination (Dpn I)	\$1.55
Ligation	\$0.32
Transformation (commercial, chemically competent cells)	\$8.60

screening protocols in this thesis is 8 days (constitutive promoters). For promoters that had to be tested with and without an inducer, the process added 10 days to the protocol. A breakdown of the time taken by intermediate steps is provided in Table 5.2. Reducing the overnight ligation step results in <10x fewer colonies post-transformation and is not recommended. Reducing the DpnI digestion is possible but not recommended to ensure that the insulated promoter library is not contaminated with uninsulated template sample.

Table 5.2: **Approximate time cost to insulate a promoter**

Protocol	Time
PCR	1 day
Dpn I restriction digest	1 day
Ligation	1 day
Transformation	1 day
Screening	2-3 days
Plasmid purification	1 day
Sequencing	1 day

The time required for ordering/synthesizing a new insulated promoter part was not be included as the time required will vary depending on the size of the insulated promoter part and the method of part creation. For example, insulated minimal

promoters ordered as as ultramer oligonucleotides and then annealed to create the cloning-ready promoter part can take as few as 3 days. However, a longer insulated promoter part that must be ordered as a gblock will take, at a very minimum, 12 days to be available for cloning.

5.3.3 Transferring the randomized spacer screening methodology to other organisms

Fluctuating promoter expression as a result of changing DNA sequence at promoter 5' junctions has been extensively studied in *E. coli*. However, the problem has been known to exist in other organisms as well (word of mouth from Prof. Karmella Haynes, Arizona State University, at the Synthetic Biology Engineering Research Center Spring Retreat 2015). Synthetic promoter parts in other host organisms may also benefit from insulation against their upstream sequences. Procedurally, there are no barriers to transferring the DNA spacer generation and screening methodology to unidirectional promoter parts in other species of bacteria, as well as in yeast and even mammalian systems. No modifications to the methodology need to be made. Insulated promoter libraries can be grown for screening using the standard growth protocols for flow cytometry for that species and type of promoter. DNA spacers can be inserted upstream eukaryotic minimal and other constitutive promoters in exactly the same way as in the bacterial constitutive promoter (See Figure 5.1). In mammalian regulated promoters the DNA spacer methodology could serve the additional purpose of adjusting promoter expression levels without having to change the locus of cloned regulator elements (such as enhancer or repressor elements), thus reducing the number of design, build test cycles. However, the spacer screening methodology cannot be applied to bidirectional promoters [55, 69]. Bidirectional

promoters are short (<1kb) regions of intergenic DNA between the 5' ends of two divergent genes on either strand of DNA. The intergenic region serves as a promoter element for both genes and adding a spacer of either side of the promoter would disrupt transcription initiation from the gene on the opposite DNA strand.

5.3.4 Generating improved context-independent variants of commonly used BioBricks promoters

As I show in the next chapter (Chapter 6, Figure 6.3) that a screened DNA spacer selected to eliminate DNA sequence context dependence of one upstream cloning scar also successfully eliminates expression variations caused by a different cloning scar. It should thus be possible to undertake the screening process for each promoter only once and generate new, improved versions of well-studied promoter parts that behave predictably regardless of variations in DNA sequence at the 5' promoter junction. I am making the insulated J23100 promoter, two insulated pBAD promoters (insulated with two distinct but equivalent DNA spacers) and one insulated pTet promoter publicly available through the iGEM BioBricks Parts Registry[35] as well as through Addgene [1].

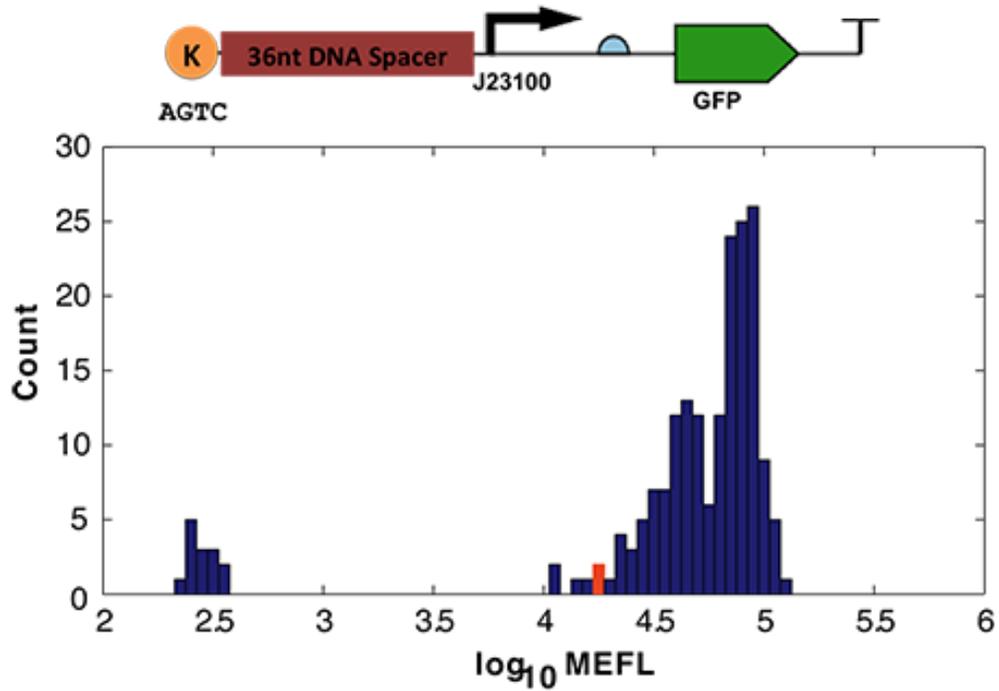


Figure 5.4: **Distribution of insulated 36nt-J23100_K GFP library.** Expression of approximately 400 samples from the 36nt-J23100_K spacer library were checked by flow cytometry for GFP expression levels. The orange bar marks the expression level of insulated J23100_K cassettes with the same expression level as the uninsulated J23100_A-GFP reference circuit. The expression distribution contained a clear peak. Samples expressing GFP at that same level as the J23100_A reference circuit fell at the lowest end of the distribution of expressions and was lower than peak insulated cassette expression. All fluorescence readouts were converted to absolute units (MEFLs).

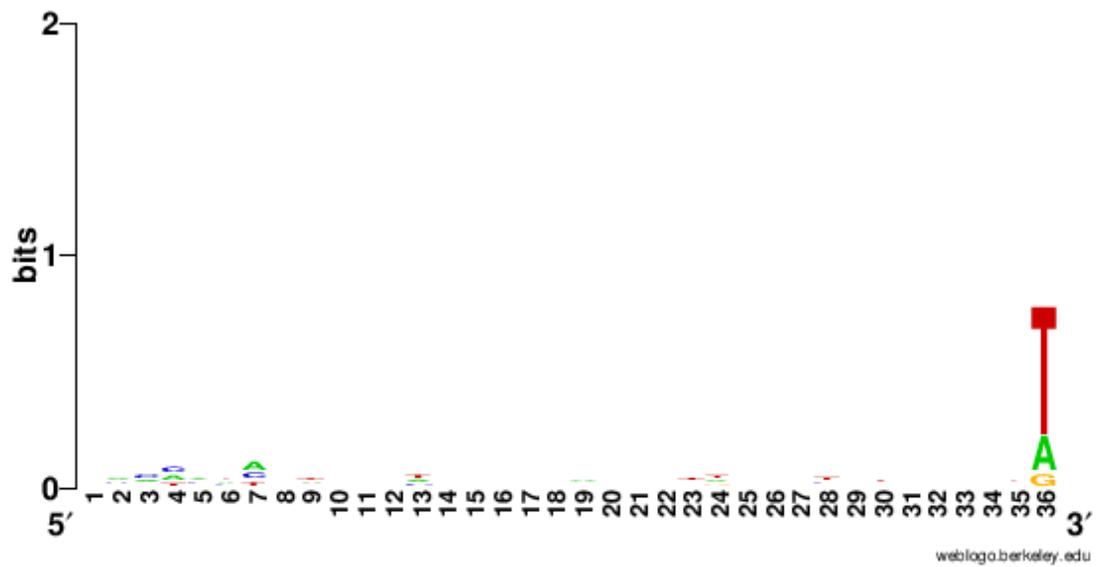


Figure 5.5: **Visualization of sequence weighting in sequenced 36nt spacers from 36n-J23100_X libraries.** 150 36nt spacer sequences from J23100_A, J23100_E and J23100_K were analyzed for consensus bases and motifs using the WebLogo 2.8.2 tool [14]. The base at the final position of all sequenced spacers was found to be (deoxy)thymine in 60% of sequenced spacers.

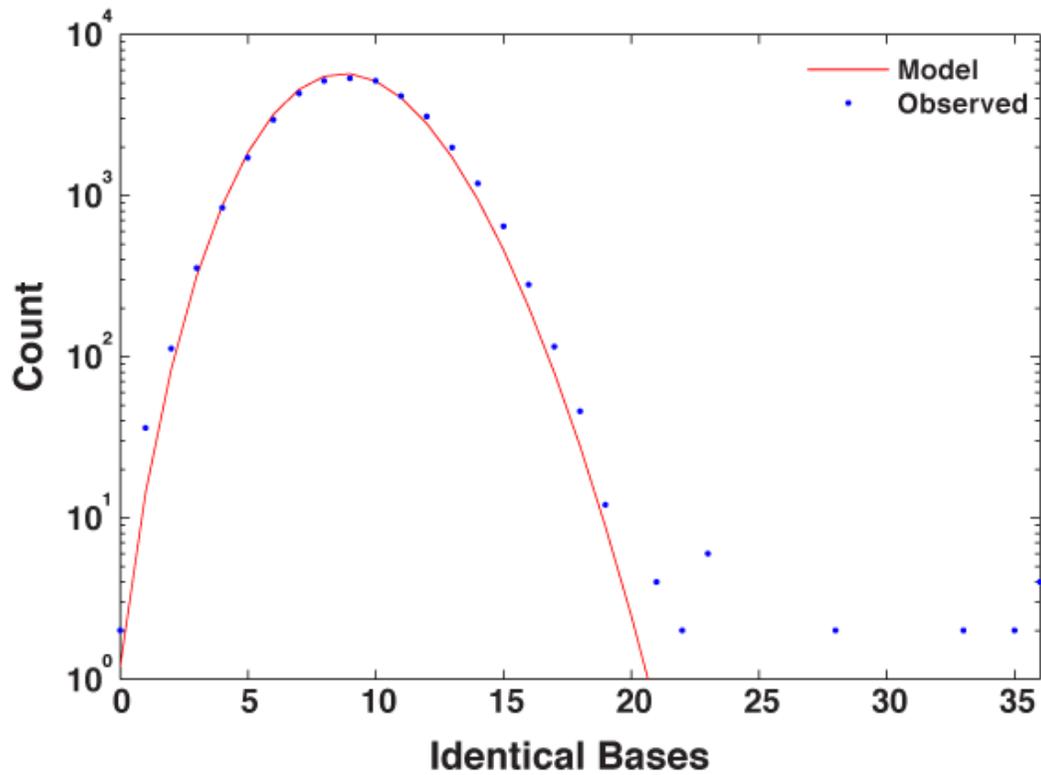


Figure 5.6: **Sequence overlap distribution for all sequenced DNA spacers.** Among the 150 36nt spacers from J23100_A, J23100_E and J23100_K plates sequenced, anomalous enriched presence of any one spacer sequence was not observed. Most sequenced spacers had approximately 9 identical bases. Only 2 duplicate spacers were found among the 150 sequenced spacers.

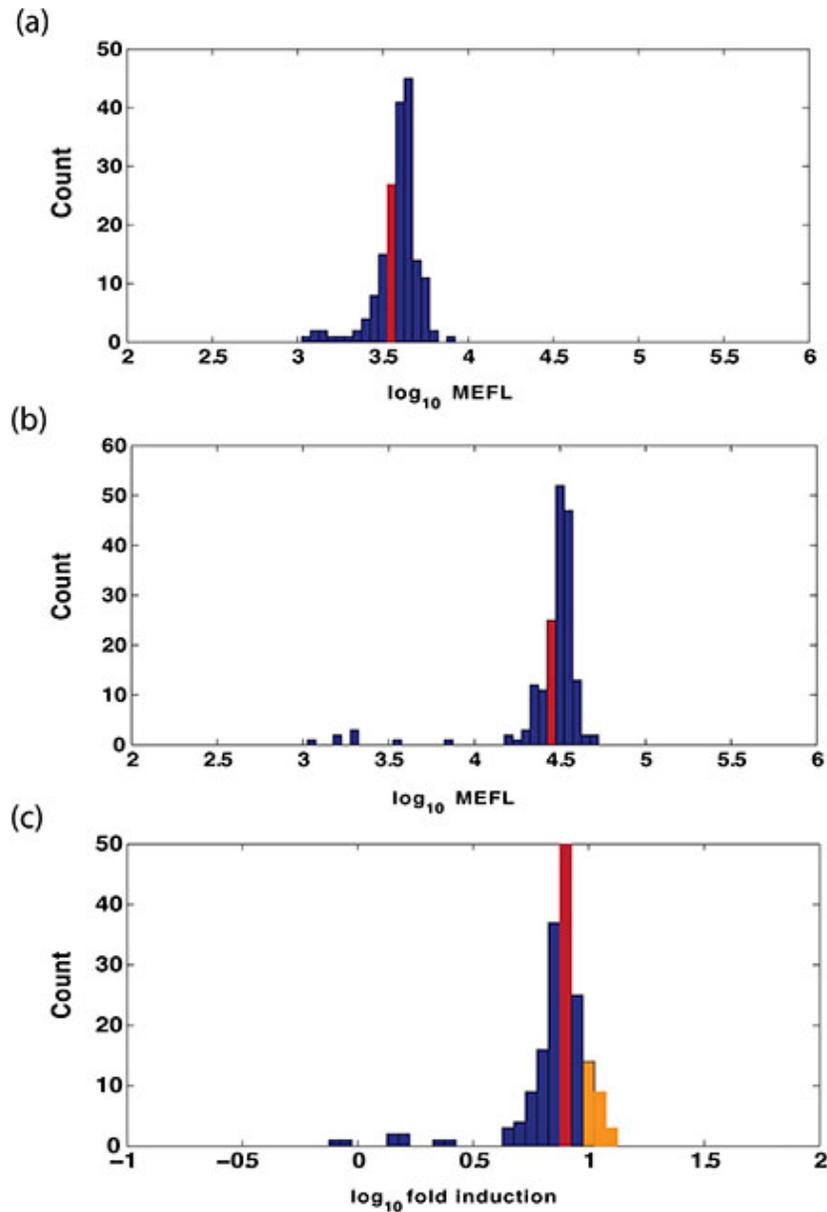


Figure 5.7: **Distribution of pBAD_F GFP library insulated with 36nt DNA spacers** Expression of approximately 200 samples from the 36nt-pBAD_F spacer library were checked by flow cytometry for GFP expression levels in the presence and absence of 1mM L-arabinose. The library was then compared to the reference device expression (red). Devices with $>10\times$ GFP expression are denoted in orange. (a) GFP expression of uninduced insulated pBAD_F-GFP circuits. (b) GFP expression of induced insulated pBAD_F-GFP circuits. (c) Log₁₀ fold change of sampled insulated pBAD_F-GFP library on induction. All fluorescence readouts were converted to absolute units (MEFLs).

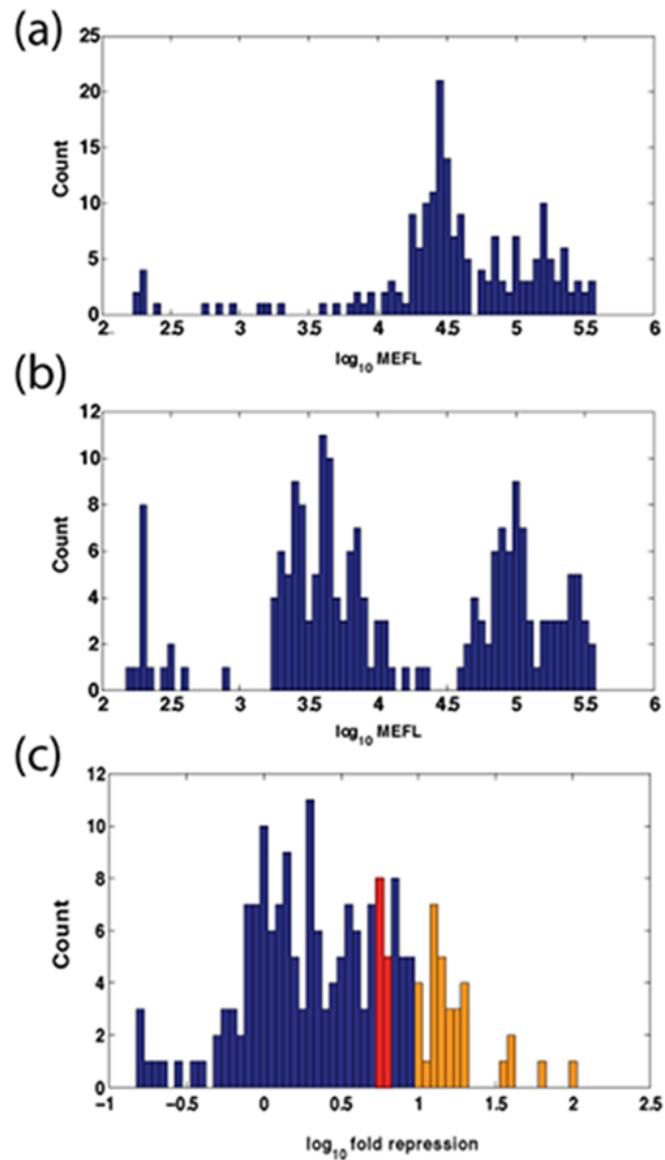


Figure 5.8: **Distribution of pTet_A RFP library insulated with 36nt DNA spacers** Expression of approximately 200 samples from the 36nt-pTet_A spacer library were checked by flow cytometry for RFP expression levels in the presence and absence of 1mM L-arabinose. The library was then screened against the reference device expression (red bar). Devices with $>10x$ RFP repression are denoted in orange. (a) RFP expression of uninduced insulated pTet_A-RFP circuits. (b) RFP expression of induced insulated pTet_A-RFP circuits. (c) Log10 fold change of sampled insulated pTet_A-RFP library on induction. All fluorescence readouts were converted to absolute units (MEFLs).

Chapter 6

DNA Spacers: Component order independence and precise expression control through the use of empirically selected DNA spacers

6.1 Introduction

In this chapter, I validate the utility of my spacer screening methodology by using the DNA spacers that I obtained through the J23100_E and pBAD_F used them to in the previously chapter and sequentially insulate the promoters within my 24 inverter permutation set (for inverter permutation circuit architecture, please see Figure 4.1). First, I show the 24 inverter set with either only J23100 or only pBAD insulated in each transcription unit permutation of the inverter (Figure 6.1 and Figure 6.2). In each case, it would be reasonable to expect that insulating one or the other affected promoter in the 24 inverter set corrects some, but not all anomalous device expression profiles (where expression either violates the NOT gate truth table, or fails to express entirely). Then, I insulate both affected promoters J23100 and pBAD at once. If my spacer screening methodology is robust, and if the majority of the anomalous expression is a result of the DNA sequence variations at the 5'

boundaries of promoters, then insulating just only J23100 and pBAD should correct anomalous expression.

Finally, selecting a DNA spacer from the pTet screen that improves the RFP repression upon inverter induction with L-arabinose, I would expect to see an improvement in the RFP fold change (greater repression) in the 24 inverter set in which all promoters are empirically insulated.

6.2 Results

6.2.1 Empirically insulating constitutive promoter J23100 in 24 inverter set improves overall device performance and reduces DNA sequence-based order dependence

Following the selection of a spacer sequence from the J23100-insulated GFP cassette screen (Figure 5.2), I replaced the J23100 promoter in all 24 permutations of my test inverter with an insulated version of J23100. The DNA spacer insulating the promoter was obtained from the set of DNA spacers that produced GFP expression levels identical to that of the uninsulated J23100_A in isolation. The inverters were assembled using MoClo DNA assembly and were grown overnight in LB Broth as described in Chapter 3. They were then grown for a second night in the presence (ON state) or absence (OFF state) or 1 mM L-arabinose. The inverters were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green and red fluorescence was measured by flow cytometry. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9]. Insulating J23100 resulted in an increase in the fraction of devices with >10x GFP and RFP induction and repression, respectively. No inverters demonstrated an anomalous pattern of expression relative

to the expected NOT-gate expression truth table. Five out of 24 of the J23100-insulated 24 inverter set failed to show any fold change.

6.2.2 Insulating inducible promoter pBAD in 24 inverter set improves overall device performance and reduces DNA sequence-based order dependence.

Following the selection of a spacer sequence from the pBAD_F insulated GFP cassette screen, we rebuilt the 24 inverter set with new insulated pBAD parts (see Figure 6.2). All instances of the pBAD promoter in each inverter were insulated. We measured the GFP and RFP fluorescence in the ON (induced, 1mM L-arabinose) and OFF (uninduced, 0 mM L-arabinose) states. Insulating pBAD resulted in an increase in the fraction of devices with >10x GFP and RFP induction and repression, respectively. No inverters demonstrated an anomalous pattern of expression relative to the expected NOT-gate expression truth table. Five out of 24 of the J23100-insulated 24 inverter set failed to show any fold change.

6.2.3 Co-insulating promoters J23100 and pBAD in the 24 inverter set eliminates upstream promoter DNA sequence-based order dependence

Figure 6.1: **Expression profile of 24 inverter set with promoter J23100 insulated empirically upon induction with 1mM L-arabinose.** GFP and RFP fold change upon L-arabinose induction of J23100-insulated 24 inverter set. DNA spacer sequence was obtained by screening the 36n-J23100_E library against J23100_A. Insulated promoters are denoted in the device key by small gray boxes placed between the purple box denoting J23100-AraC and its upstream MoClo fusion site. Insulation of J23100 resulted in a smaller fraction of both devices that fail or behave anomalously. Interestingly, out of the 8 devices that either failed or had an expression pattern violating the NOT-gate truth table, 6 devices contained pBAD_F with either GFP or TetR as the CDS. All fluorescence readouts have been converted to absolute units (MEFLs).

Figure 6.2: (Caption continued on following page.)

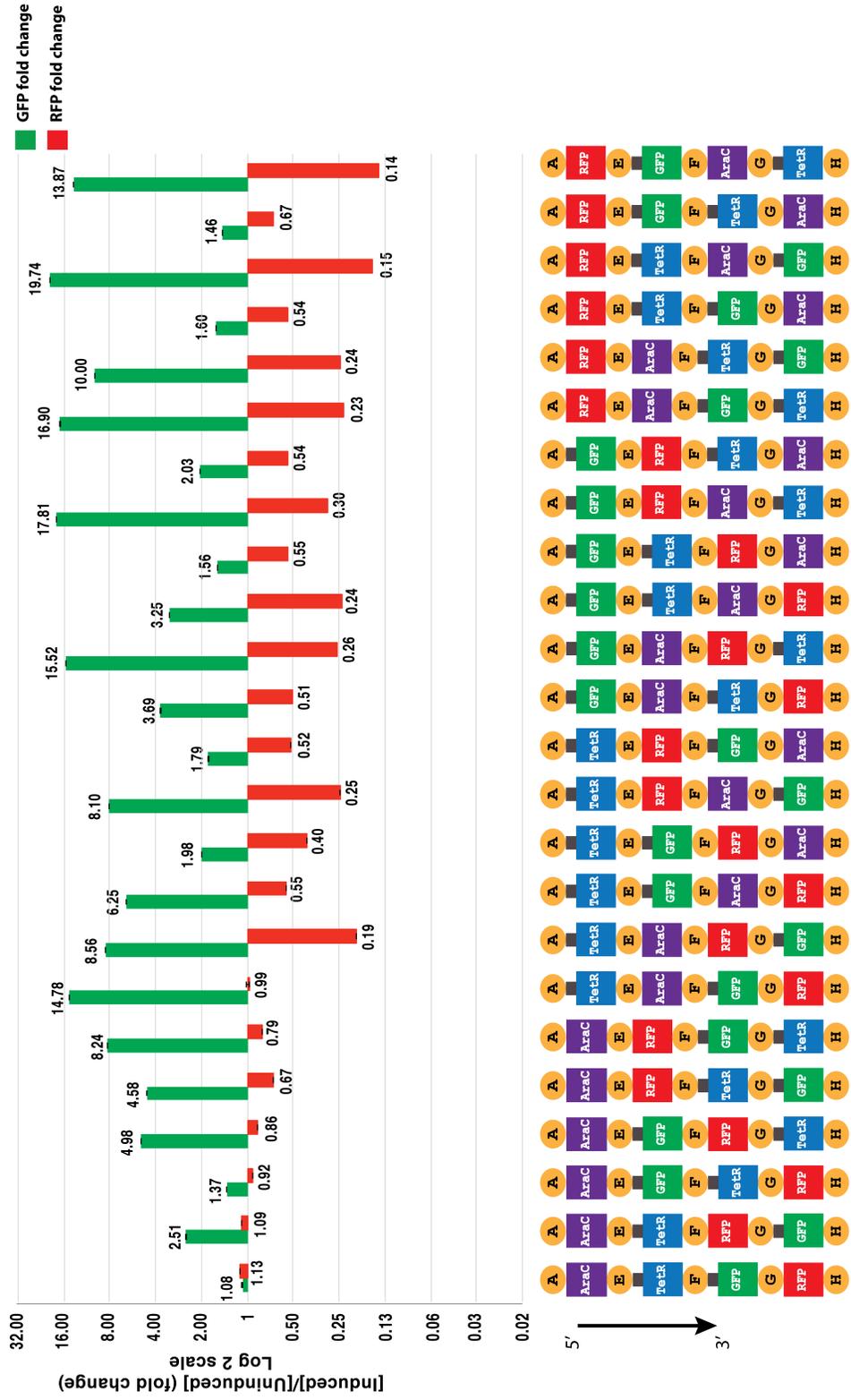


Figure 6.2: **Insulating promoter pBAD with screened DNA spacer reduces context dependence and improves overall inverter performance.** GFP and RFP fold change upon L-arabinose induction of pBAD-insulated 24 inverter set. DNA spacer sequence was obtained by screening the 36n-pBAD_F library against pBAD_A. Insulated promoters are denoted in the device key by small gray boxes placed between the upstream MoClo fusion site and the blue and green boxes denoting pBAD-TetR and pBAD-GFP TUs, respectively. Insulation of J23100 resulted in a smaller fraction of both devices that fail or behave anomalously compared to the uninsulated 24 inverter set. All fluorescence readouts have been converted to absolute units (MEFLs).

Figure 6.2.3 shows the 24 inverter set where both constitutive promoter J23100 and inducible promoter pBAD were insulated. The insulated J23100 promoter obtained from the J23100_E screen and used above (Figure 6.1). The insulated promoters used in Figure 6.2 and Figure 6.1 (obtained from the spacer screen performed on pBAD_F and J23100_E (Figure 5.7 and Figure 5.2)) were reused to insulate pBAD in the co-insulated 24 inverter set. Using these insulated promoter parts, I rebuilt the 24 inverter set using MoClo DNA assembly. All instances of promoters J23100 and pBAD in each inverter were insulated. The inverters were grown overnight in LB Broth as described in Chapter 3 and then grown for a second night in the presence (ON state) or absence (OFF state) or 1 mM L-arabinose. The inverters were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green and red fluorescence was measured by flow cytometry. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

We measured the GFP and RFP fluorescence in the ON (induced, 1mM L-arabinose) and OFF (uninduced, 0 mM L-arabinose) states. Both J23100 and pBAD promoters previously demonstrated upstream DNA sequence context dependence with MoClo fusion sites “E” and “F” for promoter J23100 and fusion site “F” for promoter pBAD. On co-insulating both promoters, all 24 devices performed as expected for a NOT-gate. Moreover, a greater fraction of inverters demonstrated >10x change in both GFP and RFP expression, resulting in better dynamic range of input and output.

6.2.4 Insulating all promoters with custom-screened DNA spacers improves overall device performance and increases the fraction of devices performing as expected

When tested in isolation, the pTet promoter was found to be robust and unaffected by its upstream DNA sequence (Figure 4.9). The 36nt-pBAD DNA spacer screen was applied to screen for spacers that would provide greater pTet repression on insulation. A DNA spacer providing 10x repression was selected and an insulated pTet promoter generated. The 24 inverter set was rebuilt with all insulated promoters from the screen. The inverters were grown overnight in LB Broth as described in Chapter 3 and then grown for a second night in the presence (ON state) or absence (OFF state) or 1 mM L-arabinose. The inverters were diluted in 1x Phosphate Buffered Saline (PBS) as described in Chapter 3, Section 3.3.2 and the expression of green and red fluorescence was measured by flow cytometry. All samples were tested in triplicate. Fluorescence measurements were converted to absolute units (MEFLs) using the TASBE Tools [7, 9].

Insulating pTet in addition to J23100 and pBAD does not alter the upstream sequence-related context independence achieved by insulating only the J23100 and pBAD promoters. Furthermore, insulating the pTet promoter increased the fraction of inverters that demonstrated >10x decrease in RFP expression on L-arabinose induction. Finally, as expected, insulating the pTet promoter did not significantly affect the GFP expression fold-change.

In a 24 inverter set with both J23100 and pBAD simultaneously insulated via our method, no devices behaved anomalously (Fig. 6.4) and the fraction of inverters showing a 10x fold change for both GFP and RFP also increased. Using an insulated pTet promoter from our screen in conjunction with J23100 and pBAD did

not cause anomalous behavior of any inverters (Fig. 6.4). Furthermore, as expected, insulating pTet improved RFP fold change (Fig. 6.3) without impacting GFP fold change, as candidate pTet spacers with off target effects were eliminated through the spacer screening process. Finally, the 24 inverter set with all promoters insulated demonstrate the highest number of devices with $> 10x$ RFP repression on arabinose induction.

Figure 6.4 demonstrates the observed improvement of GFP expression performance as a function of the number of empirically-insulated promoters included in the inverter design. For uninsulated inverters (pink), only 10% of the 24 inverters showed in five instances a 10x induction in GFP expression on L-arabinose induction. In inverters where only either promoter J23100 (blue) or pBAD (green) was insulated, under 30% and under 40% of inverters were observed to show a 10x increase in GFP expression upon L-arabinose induction. In contrast, in inverters where both promoters J23100 and pBAD were insulated simultaneously (red), approximately 50% of all inverters could be showed a 10x increase in GFP expression. This number increased, but not substantially, when all promoters were insulated (black), which is not surprising as promoter pTet controlled RFP expression and insulating promoter pTet should not impact GFP expression significantly.

Figure 6.4 demonstrates the observed RFP repression profile improvement as a function of the number of empirically-insulated promoters included in the inverter design. No uninsulated inverters (pink) are expected to show a 10x repression of RFP expression on L-arabinose induction. Furthermore, almost 60% of all uninsulated inverters are expected to show no RFP repression upon induction. In contrast, 100% of all-insulated (black) or promoters J23100 and pBAD simultaneously insulated (red) inverters are expected to show some RFP repression when inverter circuits

are insulated. A greater area under the graph represents better overall performance. Whereas all-insulated (black) and J23100- and pBAD-only (red) insulated sets contained near-identical areas under their respective curves in the GFP expression profile (Figure 6.4), the all-insulated set (black) in the RFP performance profile contains more area under the curve than that under the J23100 and pBAD-only (red) curve, demonstrating that insulating the pTet promoter improved the pTet performance specifically without impacting the GFP expression profile.

6.3 Discussion

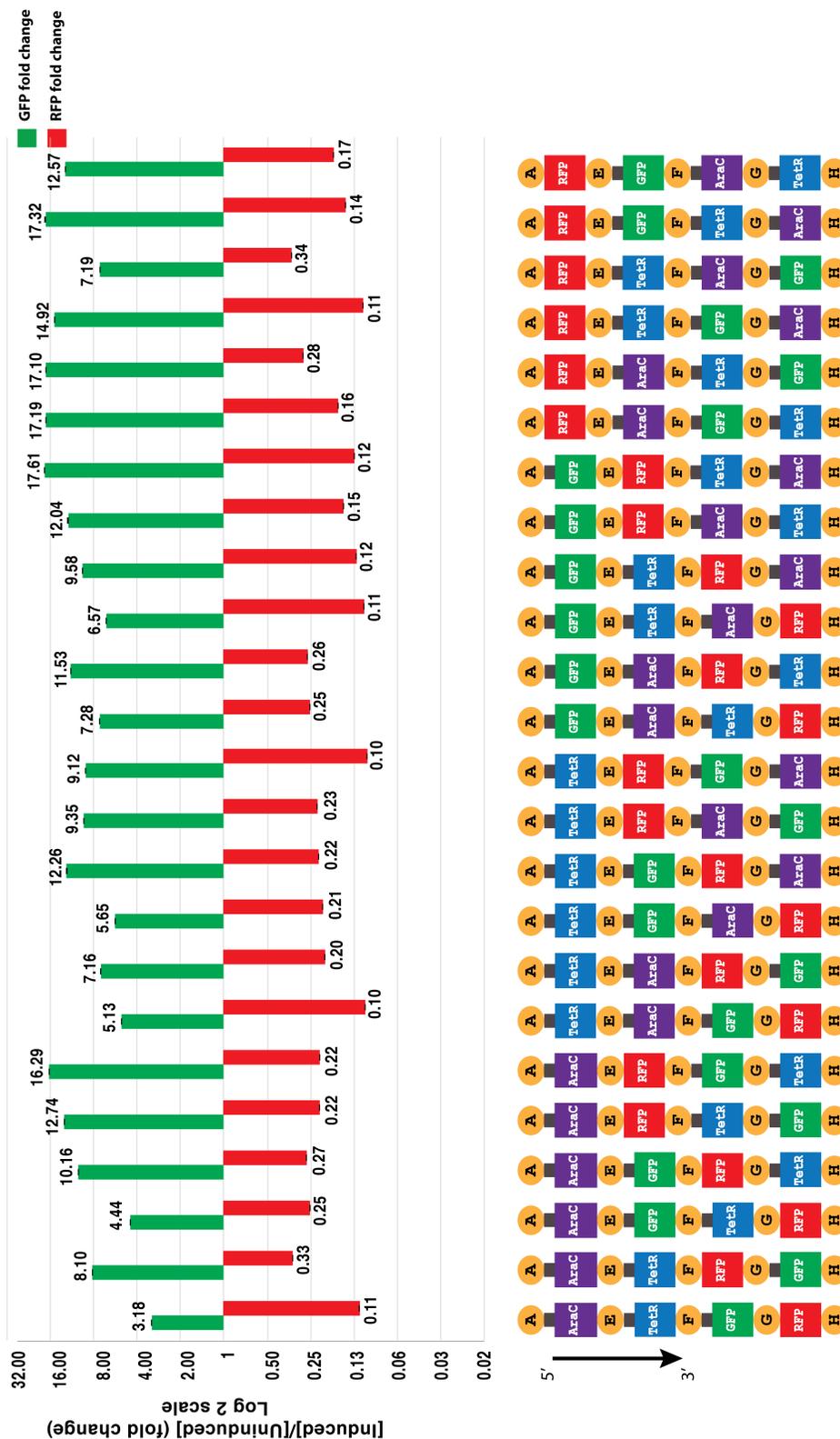
Part-junction interference at the 5' boundaries of promoters is unavoidable if they are to be composed into complex devices. Even small variations in the DNA sequence at the 5' boundaries of promoters can greatly alter promoter performance leading to design failures of composed genetic circuits. There does not appear to be any standardized set of rules for selecting DNA spacer sequences that provide reliable insulation without itself affecting promoter expression. We provide a method for generating and screening a library of insulated promoters within the context of the target device for precise expression control that we show to work for constitutive, inducible and repressible promoters.

Given the very large range of expression levels observed in the sampled J23100 and pTet insulated promoter libraries, it is unlikely that designing a DNA spacer that provides reliable insulation without itself affecting promoter expression can be achieved without extensive testing. A DNA spacer that reliably insulates one promoter may not work when inserted upstream of a different promoter. Given the ability of our method to generate and screen in the context of the target device, our method has the potential to be used as a tool for fine-tuning circuit expression

without having to swap out promoters and RBSs. It could also potentially be used to increase the separation of reporter expressions between ON and OFF states in binary logic devices.

The distribution of expression for screened insulated pBAD promoters (Figure 5.7) was found to be much tighter than that for promoters J23100 and pTet (Figures 5.2 and 5.8). It is possible that for a tested promoter, the distribution of insulated promoter expression will not span the required expression range needed. In such a case, swapping promoter/RBS parts may become unavoidable. In every case here, we used a reference device to screen candidate insulated promoter circuits against. In the absence of a reference, our method has no level to match, but can still achieve precise expression control, but would still be useful for eliminating candidate DNA spacers that abolish expression or cause unanticipated off-target effects. Applying our method to genetic circuits in which component modules are on different plasmids requires a more laborious and circuitous route than we have outlined, and has not yet been tested.

Figure 6.3: (Caption continued on following page.)



Expression profile of the 24 inverter set with promoters J23100 and pBAD insulated empirically upon induction with 1mM L-arabinose. GFP and RFP fold change upon L-arabinose induction of 24 inverter set in which promoter J23100 and promoter pBAD are insulated. DNA spacer sequences were obtained from the 36n-J23100_E and 36n-pBAD_F screens. Insulated promoters are denoted in the device key by a small gray box placed between the purple, blue or green box denoting J23100-AraC, pBAD-TetR and pBAD-GFP, respectively, and its upstream MoClo fusion site. Simultaneous insulation of known affected promoters completely eliminates genetic context-based order dependence.

Figure 6.3: Expression profile of the 24 inverter set with promoter all promoters insulated empirically upon induction with 1mM L-arabinose. (Caption continued on following page.)

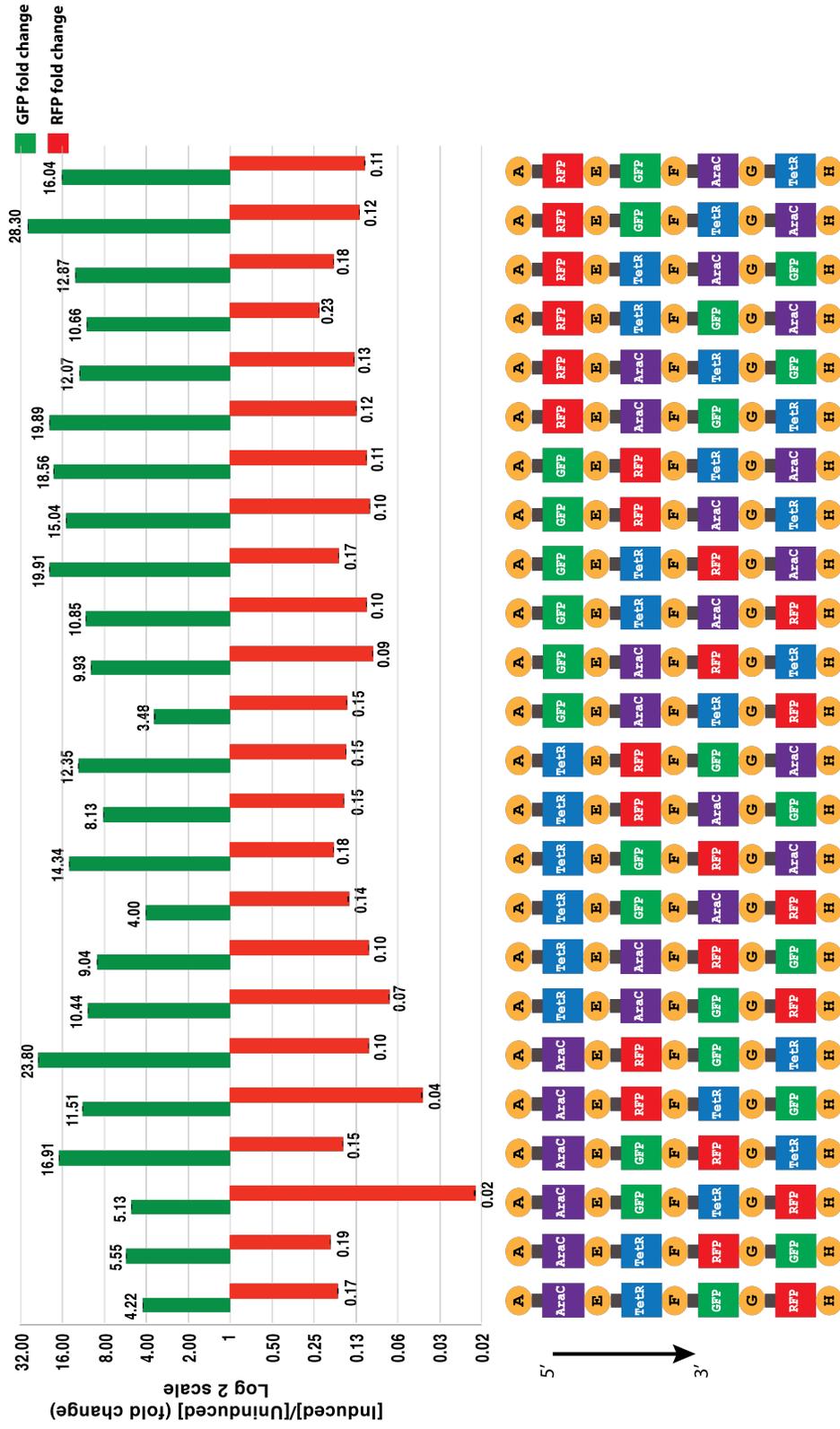
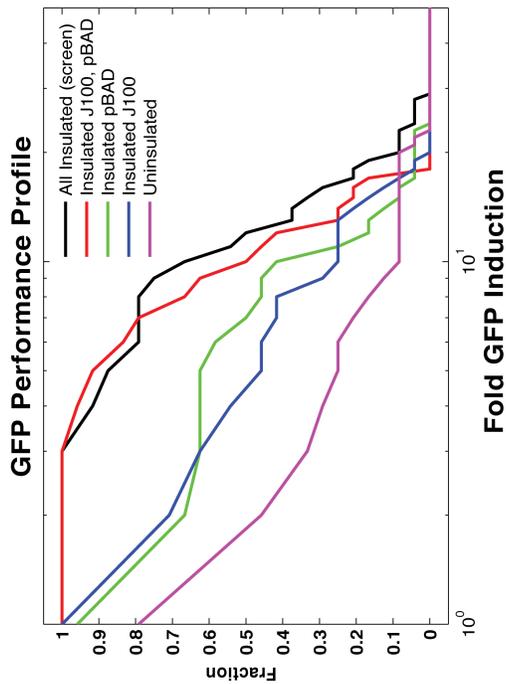


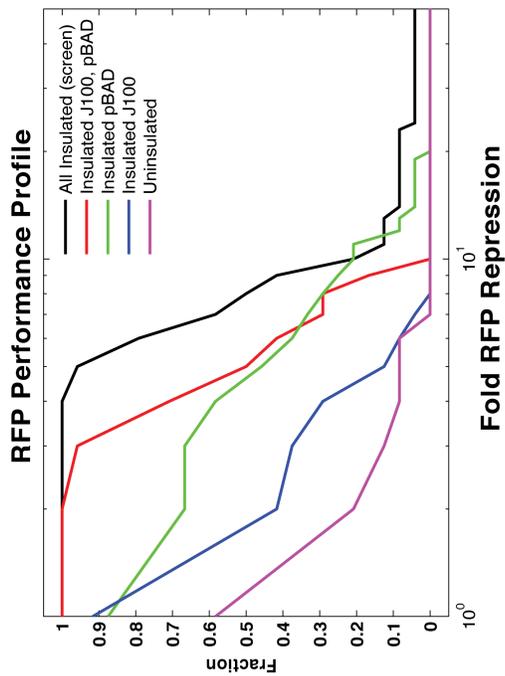
Figure 6.3: **Expression profile of the 24 inverter set with promoter all promoters insulated empirically upon induction with 1mM L-arabinose.** (Continued from previous page.) Insulating the upstream DNA context-independent promoter pTet improves inverter performance. GFP and RFP fold change upon L-arabinose induction of J23100, pBAD and pTet-insulated 24 inverter set. DNA spacer sequences were obtained from the 36n-J23100_E, 36n-pBAD_F and 36n-pTet_A screens. Insulated promoters are denoted in the device key by a small gray box placed between the purple, blue, green or red box denoting J23100-AraC, pBAD-TetR, pBAD-GFP or pTet-RFP, respectively, and its upstream MoClo fusion site. Selecting a DNA spacer from 36nt-pTet screen that provided a >10x repression on L-arabinose induction improves the RFP fold change in the 24 inverter set.

Figure 6.4: **Observed incremental improvement in GFP and RFP expression of the 24 inverter set upon insulating with DNA spacers selected using screening methodology.** Observed fraction of inverters with a 10x GFP increase in GFP expression (a) and 10x decrease in RFP expression (b) upon L-arabinose induction increases with the number of promoters insulated using our screening technique. Pink line denotes uninsulated inverters. Blue and green lines represent J23100-only insulated and pBAD-only insulated 24 inverter sets, respectively. Red line represents 24 inverter set with promoters J23100 and pBAD simultaneously insulated. Black line represents 24 inverter set where all promoters are insulated.

(a)



(b)



Chapter 7

DNA Assembly Automation:

Development and validation of liquid classes for automated DNA assembly on a Tecan liquid-handling robotic platform

7.1 Introduction

As they grow in size and complexity, designing, building, organizing, tracking and managing genetic devices and device-variants manually becomes both error-prone and time-prohibitive. A four-part expression module consisting of a promoter, RBS, Coding Sequence (CDS) and terminator and lacking any logical computation and built from a modest part library comprising 25 promoters, 5 RBSs, 5 CDSs and 1 terminator has 625 possible valid devices. When engineering genetic devices comprising multiple intermediate modules, additional variations involving orientation of promoters relative to each other further increases the size of the design space. As such, manually assembling DNA is impractical, inefficient, and highly error-prone. Standardized and efficient, multi-part DNA assembly methods [68, 38, 20, 44, 19, 63] combined with computational DNA assembly planning tools [6] can help make DNA assembly more efficient, but even multiplexing-friendly DNA assembly methods [38]

cannot be solely managed manually. Commercial DNA synthesis presents an alternative to manual DNA assembly in the lab, but the cost of DNA synthesis is still prohibitive as the only DNA synthesis strategy is too high to replace manual DNA assembly in research labs. As such DNA assembly is often the rate-limiting step in synthetic genetic workflows and would greatly benefit from being automated.

7.1.1 Existing automation in synthetic biology

Automation in synthetic biology is not a novel idea. Computational tools automating individual steps in the specify-design-build-test workflow already exist [66, 12], and new tools are continuously being added and existing tools being improved upon. Synthetic biology data management frameworks such as Clotho [70] are also being developed as platforms providing an end-to-end automated processing of all specification and design steps in a synthetic biology workflow allowing for seamless management of bioengineering design and analysis. Other end-to-end workflows like the Tool-Chain to Accelerate Synthetic Biology Engineering (TASBE) [8] also attempt to automate the engineering workflow of biology. The biological design specification language EUGENE automatically generates the possible biological designs for assembly from a user-defined library of genetic parts and applies any user-desired assembly constraints to the design space [53, 54, 10, 17].

The physical assembly of DNA parts has remained, since the beginning, a major broken link in synthetic biology research automation. Multiple automated tools exist for most other steps in the synthetic biology workflow: A formally expression logic function can be automatically converted into biological designs using defined parts and part features [42]. These parts can be automatically mapped to characterized physical DNA samples stored in part repositories. Target circuits and genetic devices

can be automatically created by mapping these parts and features on to DNA backbones to create plasmids, which are then stored in sample tracking and management systems [60, 15] and can be processed in bulk automatically. Tools exist for generating automated assembly plans for constructing devices to generate the physical samples for the devices that the design abstraction tools have built, and even provide users their choice of one or a combination of the most-used DNA assembly methods [6]. Once the samples are generated – mostly manually – they can be tested using high-throughput samplers attached to flow cytometers. Their data can be automatically processed and analyzed [8]. New machine-learning tools are even available to automatically rank the collected data from the devices using user-defined ranking parameters [61]. Furthermore, tools exist for automatically generating physical representations of standards-compliant design visualizations [24].

7.1.2 Current state of DNA assembly automation

Until the advent of synthetic biology and biological part repositories such as Joint BioEnergy Institute (JBEI)-ICE Registry [31], the CIDAR-ICE repository [56] and the public genetic parts repository, Registry of Standard Biological Parts maintained by the International Genetically Engineered Machine (iGEM) Foundation [35], genetic parts were not stored in any standardized format. Only a small handful of parts needed to be cloned per project, and thus, the volume of cloning done was not very large, these were not big hurdles in the realization of actual research aims. Molecular cloning was always done *ad hoc* by performing restriction digests using whatever restriction endonuclease sites were available flanking the part of choice. The lack of standardization made it impractical to automate any cloning protocol as it was unlikely that the next cloning reaction would follow an identical protocol.

Furthermore, automating DNA assembly required significant monetary investment in equipment (liquid handling robots and their associated paraphernalia) as well as the time investment for the generation of instructions to develop accurate parameters for handling the various classes of liquids being processed and the precise series of instructions to run these protocols. It would be impractical to make these investments when DNA assembly was not the bottleneck in the research pipeline. The mindset that DNA assembly is an achievement of intensive labor rather than efficiency remained ingrained even after the advent of synthetic biology and even researchers have been much slower in recognizing the pressing need for automation of DNA assembly compared to other steps in the synthetic biology workflow. As a result, DNA assembly automation lags behind the automation of other steps in the synthetic biology pipeline in spite of often being the bottleneck in the number of synthetic devices being built and tested.

Furthermore, not all steps in molecular cloning are easily automatable: selecting and picking clones from a transformation plate post-cloning involves many complex decisions that take a human with fundamental knowledge in molecular biology and basic lab training a matter of seconds to minutes, but involve too many complex decision-making processes that cannot be taught to a machine at the current state of the art. However, sample-tracking and liquid-handling steps are easily automatable and also constitute a major source of human error in manual DNA assembly

Nonetheless, partial automation of laboratory protocols that are essential to carrying out DNA assembly have been developed: Qiagen, a leading vendor of DNA extraction and purification kits, also sells its “off-the-shelf” bench top robot for executing only its own protocols. However, it processes only 12 samples at a time, is not customizable, and is slower than manual execution of these protocols. Ven-

dors of liquid-handling robotic systems have also developed protocols for carrying out DNA extraction and purification using their robotic technology. While these are much higher-throughput and also fast, they are proprietary and are only available as supplementary purchases over and above the price of the liquid-handling robot itself.

7.2 Optimizable Liquid Class Parameters

Section 7.3 provides comprehensive listings of all the optimized liquid class parameters that I made to increase the automated MoClo DNA assembly reaction efficiency from 0% to approximately 98%. To understand the changes to the liquid classes I have made the following list, obtained from the Tecan [37] EVO 150 reference manual, provides a description of the function of each liquid class parameter. Parameter descriptions listed below in Section 7.2 are either paraphrased or reproduced from Tecan EvoWare user guide. [37].

7.2.1 Liquid Class Parameters

Aspiration Speed: Speed at which liquid is drawn into the pipette arm. Aspiration speed is set in $\mu\text{L}/\text{s}$. A lower aspiration speed can improve the pipetting accuracy, especially for viscous liquids. A higher aspiration speed improves the speed of operation.

Delay: Specifies the time the pipette is held in aspirate position before being retracted to “Retract to” position. Specified in ms.

System Trailing Airgap (STAG): Specifies the volume of air separating the system liquid from the sample being aspirated and dispensed.

Leading Airgap (LAG): The air in the leading airgap is dispensed together with the net sample volume in the *Dispense* command. The resulting tip blowout leads to more accurate pipetting. The LAG creating additional buffering between the sample and the system liquid although that is not its primary purpose. Specified in μL .

Trailing Airgap (TAG): The trailing air gap specifies an additional, user-defined volume of air to be aspirated into the pipette tip after aspirating the specified sample volume. The TAG is used to prevent liquid dripping from the tip when the robotic arm moves. In addition, it ensures that the liquid meniscus is not dragged away from the end of the tip during fast movements. Specified in μL . The TAG is aspirated at the Z-position specified with the *Retract Tips to*. A Z-position relative to the liquid level requires liquid level detection and must not be used when carried out with septum-piercing as the septum interferes with the capacitance-based liquid detection system and results in TAG being aspirated while the tips are still in the liquid.

Excess Volume: Used to specify excess volume of sample aspirated in addition to the volume specified in the experiment parameters. The purpose of the excess volume is to reduce the contamination of the sample liquid in the tips with the system liquid to an absolute minimum. If excess volume is specified, a destination for discarding excess volume must also be specified (for example, *Discard to waste*, or *Back to vessel*). Specified in μL or as %

Conditioning Volume: The conditioning volume can be used to maximize pipetting accuracy by ensuring that the liquid meniscus at the end of the tips is the same shape as after aspirating as it will be after dispensing. The conditioning volume is discarded directly after aspirating. If a conditioning volume is specified, a destination must also be specified for it to be discarded (for example, *Discard to waste*, or *Back to vessel*). Specified in μL or as %

Use Pinch Valve: This is used to activate the Low Volume option if it is fitted to the pipetting instrument. Works with only the first pipette tip.

Use Liquid Detection: This activates liquid level detection using a capacitive liquid level detection (cLLD) when aspirating. cLLD detects the surface through a change in the electrical capacitance of the pipetting tip when the tip enters the liquid. It is carried out just before aspirating and can be used to minimize tip contamination by reducing the tip immersion depth. A minimum volume of 30 μL in excess of the volume to be used in the reaction is required in each well for liquid level detection for water, enzyme buffers and DNA. For viscous liquids like enzymes, an even larger volume of liquid is needed to account for the liquid loss due during pipetting steps. This volume is referred to as “dead volume”. cLLD does not work well with liquid samples that do not contain adequate dissolved solutes, such as deionized water. Using cLLD with deionized water produces a non-fatal error during protocol execution.

On Detection Error: Specifies the error-handling if Freedom EVOware detects no liquid or not enough liquid. The error-handling mode *User prompt* shows the Liquid Detection Error dialog; the other 3 options carry out the chosen error-handling

action directly and do not show the Liquid Detection Error dialog.

Use Exit Signal Detection: Activates capacitive exit signal detection (dip out signal detection). This detects the liquid surface through a change in the electrical capacitance of the pipetting tip when the tip is retracted and exits the liquid. It is carried out just after aspirating. The measured liquid level is then compared with the liquid level that should have resulted from the aspiration. This can only be activated if liquid level detection has been enabled.

On Exit Signal Error: Specifies the selected required error-handling if the exit signal error does not take place as expected when retracting the tips. Currently reported in the report file from the *Export Data* command.

On PMP Clot Error: Allows selection of the required error-handling if a clot is detected using the pressure-based clot-detection system, which requires the PMP hardware option available only for Liquid-Handling (LiHa) Disposable Tips (DiTi).

Aspiration position: Specified by the Z-position and the Z-offset of the tips when aspirating. A positive value for the offset lowers the tips. If liquid level detection is being used, only z-positions relative to liquid level can be specified. X- and Y-positions can also be specified when aspirating.

Mix before Aspiration: This option specifies that Freedom EVOware is to mix the liquid before aspirating. The number of mix cycles can range between 1 and 99 and the mix volume can range between 1 and 2500 μL . The Z-position and Z-offset

can be specified by the user. A positive value for the offset lowers the tips.

Retract Tips to: Specifies the Z-position and the Z-offset to which tips should retract (move upwards) to after.

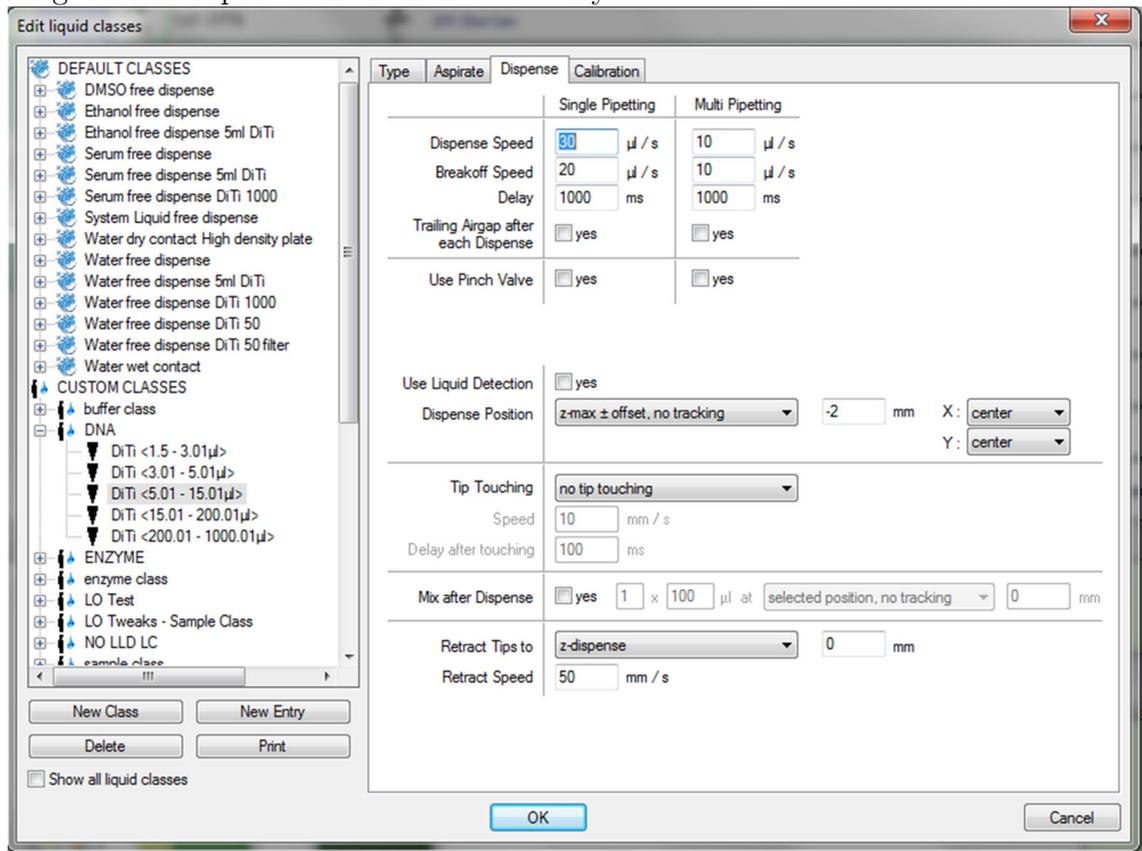
7.2.1.1 Adjustable parameters in Tecan Evo Liquid-handling robot Liquid Class Dispense protocol:

The following are the available parameters of movement and air-displacement of the disposable pipette tip that can be customized executing dispense parameters in an air-displacement Tecan liquid-handling robot. Figure 7.1 provides a screenshot of the dispense parameter setup screen within the Freedom EVOware software. The following parameter descriptions were either paraphrased or reproduced from the the Tecan EVOware user guide [37].

Dispense Speed: Specifies the speed at which aspirated sample liquid is released into the destination well. A high dispense speed is needed to achieve a proper free dispense. However, the dispense speed must be lowered when handling liquids of high viscosity. the maximum dispense speed is about 80% of the syringe size (dilutor capacity) per second. The dispense speed is specified as $\mu\text{L/s}$

Breakoff Speed: Specifies the final destination speed (breakoff) of the sample liquid from the pipette tip. The breakoff speed should normally be about 70% of the actual dispense speed but should not exceed 45% of the syringe size (dilutor capacity) per second. A high breakoff speed is necessary to achieve a proper free dispense. The

Figure 7.1: Screenshot of Dispense liquid class optimization panel in the Tecan EVOware software. Parameters must be adjusted for very-small ($1.5 - 3.0\mu\text{L}$), small ($3.0 - 5.0$ and $5.0 - 15.0\mu\text{L}$), mid-ranges ($15.0 - 200\mu\text{L}$) and large ($>200\mu\text{L}$) volume ranges for all liquid with dissimilar viscosity.



break off speed is specified as $\mu\text{L}/\text{s}$.

Delay: Specifies the time for which the pipette tip remains in the dispense position before retracting to the *Retract Tips to* position. Specified in ms.

Trailing Airgap after each Dispense: Allows for automatic aspiration of a Trailing Airgap (TAG) after each dispense. These parameters are specified in the *Aspirate* dialog box.

Use Pinch Valve: This is used to activate the Low Volume option if it is fitted to the pipetting instrument. Works with only the first pipette tip.

Use Liquid Detection: This activates liquid level detection using a capacitative liquid level detection (cLLD) when dispensing. cLLD detects the surface through a change in the electrical capacitance of the pipetting tip when the tip enters the liquid. cLLD when dispensing can be used to ensure that the pipette tip remains immersed in liquid during dispensing and is a useful parameter for dispensing very small volumes of liquids.

Dispense Position: Specifies the Z-position and the Z-offset of the tips when dispensing. A positive value for the offset lowers tips. If *Liquid Level Detection* is activated, the Dispense Position has to be specified relative to liquid level. If tracking is selected, the tips will move upwards as the liquid level rises in the wells during dispensing. This option minimizes contamination by restricting the depth to which the pipette tip is immersed into the sample.

Tip Touching: Activating *Tip Touching* during the *Dispense* command dispenses the liquid in the normal way, but then moves the tip to the side of the well and back before retracting them, mimicking manual dispense actions. This helps remove any droplets of liquid which may be adhering to the tips. This is effective when dispensing small volumes of liquid into an empty destination well. The direction of *Tip Touching* can be specified.

Speed: Specifies the *Tip Touching* speed in mm/s

Delay after touching: Specifies the delay between tip touching the well and moving back to the middle of the well.

Mix after Dispense: Allows the user to mix the pipetted sample after dispensing it. The number of mix cycles (range: 1 to 99) and the mix volume (range: 1 to 2500 μL) can be specified.

Retract Tips to: Specifies the Z-position and the Z-offset to which the tips should retract (move upwards) after dispensing. A positive value for the offset lowers the tips. If Liquid Level Detection has been activated, then only positions relative to liquid level can be specified.

Retract Speed: Specifies the required retract speed in mm/s.

7.2.2 Global Liquid Class Settings

The following settings apply to all liquid classes and pipetting steps for the Tecan Evo 150:

- First pipette tip: The first pipette tip on the Tecan Evo 150 is never used for any reaction setups.
- Pinch Valve Set to Off: This affects only the first pipette tip so is set to be off for all liquid classes and subclasses.

7.3 Liquid Class Optimization

Liquid class optimization for automated assembly of DNA parts is achieved by serially adjusting available parameters in *Aspirate* and *Dispense* liquid class protocols. The effectiveness of each adjustment in the enzyme liquid 2.0 - 3.5 μL sub-class was measured by calculating the number of true positive colonies per 100 apparent positive colonies obtained from blue-white screening of a 20 μL Modular Cloning (MoClo) [38] DNA assembly reaction creating a constitutively expressed GFP cassette. The strong constitutive promoter J23100, RBS B0032m, GFP gene and double terminator B0015 DNA parts are inserted into the Kanamycin-resistant destination vector DVL1_AE. Each parameter test was repeated 8 times.

True positive colonies were calculated by first selecting white colonies from blue-white screening from a MoClo DNA assembly transformation plate. White colonies were then both checked for color under UV light followed by DNA sequence verification. Liquid class optimization of other enzyme sub-classes and the DNA liquid class were achieved by checking the precision and accuracy of the *Aspirate* and *Dispense* protocols for liquid volumes within the sub-class manually by hand-pipetting. Each optimization step was repeated 16 times. The efficiency of our protocol (>94%), as measured by the ratio between positive colony forming units (CFUs) to total number of CFUs, is equal to that of MoClo reactions prepared manually.

7.3.0.1 Optimized liquid class parameters

Enzymes are generally supplied as viscous liquids often containing 50% glycerol. Optimizing liquid class parameters for robotically aspirating and dispensing very small volumes (2 μL) of enzymes was thus especially challenging. However, we found that the optimized liquid class for DNA is also sufficient for robotic pipetting of both

water and enzyme (ligase) buffer. To simplify our protocols, we used the same liquid class for water, DNA and buffer in our automated DNA assembly protocols. All parameters were adjusted for single-pipetting only to minimize the risk of contamination. The optimized liquid class parameters for enzyme aspirate and dispense are listed in Tables 7.1 and 7.3, respectively.

7.3.0.2 Enzyme Liquid Class: Aspirate Protocol Parameters

Liquid Density: 1.131 mg/ μ L

Table 7.1: Aspirate parameters for enzyme liquid class by volume-based sub-classes

PARAMETER	LIQUID SUB-CLASS		
	2 - 3.5 μ L	3.51 - 10 μ L	10.1 - 200 μ L
Speed	2 μ L/s	5 μ L/s	50 μ L/s
Delay	1000 ms	1000 ms	1000 ms
System Trailing Airgap (STAG)	20 μ L	20 μ L	20 μ L
Leading Airgap (LAG)	10 μ L	15 μ L	15 μ L
Trailing Airgap (TAG)	0 μ L	1 μ L	1 μ L
Excess Volume	0 μ L	0 μ L	0 μ L
Conditioning Volume	0 μ L	0 μ L	0 μ L
Pinch Valve	Off	Off	Off
Liquid Level Detection	Off	Off	Off
Mix before Aspiration	No	No	No
Aspirate Position	z-max with -1mm offset (no tracking)	z-max with -1 mm offset (no tracking)	z-max with -1mm offset (no tracking)
Retract Tips to	current position - 5 mm	z-dispense - 5 mm	current position - 5 mm
Retraction Speed	5 mm/s	5 mm/s	20 mm/s

Modified enzyme liquid class aspirate parameters improving MoClo reaction efficiency for 2-3.5 μ L subclass

Table 7.2: Stepwise optimization for aspirate protocol for enzyme liquid class. ¹Numbers are mean of 8 trials per adjusted parameter. Modified dispense protocol parameters as found in Table 7.3 were used to dispense samples while testing aspirate parameter in Table 7.1.

Modified Aspirate Parameter	# blue colonies (negative)	# white colonies (positive)	# true positive/100 white colonies
Aspirate Speed	>1000	1	1
Delay	>1000	3	3
Aspirate speed + Delay	476	10	9
Aspirate speed + Delay + Aspirate Position	5	312	100

Aspirate Speed: Reducing the aspirate speed from 5 μ L to 2 μ L allows sufficient time for the viscous liquid to get pipetted into the tip.

Delay: Doubling the delay time (500 ms to 1000 ms) during the aspirate step, and increasing the dispense delay by 10x (from 200 ms to 2000 ms) allowed sufficient time for enzymes (50% glycerol) to be aspirated completely resulting in increased MoClo efficiency. When the delay parameters of both the aspirate and dispense steps were increased, MoClo reaction efficiency went up from <1% to ~5%.

Aspirate Position: The quantity of enzyme added to the MoClo reactions is critical to the overall reaction efficiency. Aspirating from the bottom of the well ensures that the cloning reaction mix receives a sufficient quantity of enzymes resulting in an increase in MoClo reaction efficiency from ~5% to ~100%.

7.3.0.3 Enzyme Liquid Class: Dispense Protocol Parameters

Liquid Density: 1.131 mg/ μ L

Table 7.3: Dispense parameters for enzyme liquid class by volume-based sub-classes

PARAMETER	LIQUID SUB-CLASS		
	2 - 3.5 μ L	3.51 - 10 μ L	10.1 - 200 μ L
Speed	5 μ L/s	5 μ L/s	100 μ L/s
Breakoff speed	10 μ L/s	10 μ L/s	100 μ L/s
Delay	2000 ms	2000 ms	2000 ms
Trailing airgap after each dispense	No	No	No
Pinch Valve	No	No	No
Liquid Level Detection	Yes	Yes	Yes
Dispense Position	Liquid level detection +1 mm offset (center)	Liquid level dection +1 mm offset (center)	Liquid level dection +1 mm offset (center)
Tip Touching	No	No	No
Mix after dispense	No	No	No
Retract Tips to	Current position -5 mm	Current position	Current position
Retract Speed	50 mm/s	50 mm/s	50 mm/s

Modified enzyme liquid class dispense parameters that improved accuracy and precision of dispense protocol for 2-3.5 μ L subclass

The default, unmodified dispense protocol parameters yielded no positive clones in MoClo reactions. Dispense protocol parameters were modified prior to making aspirate protocol parameter improvements using 50% glycerol followed by manually

measuring accuracy and precision of dispensed liquid in a series of trials. MoClo reactions performed using the most precise and accurate dispense protocol parameters, but with default aspirate protocol parameters, resulted in no positive clones. However, using the modified aspirate protocol with the modified dispense protocol also positive clones in MoClo reactions as found in Table 1.3.

Dispense Speed: Slowing down the dispense step to 0.0125x of the default dispense speed (from 400 $\mu\text{L}/\text{s}$ to 5 $\mu\text{L}/\text{s}$) resulted in MoClo reaction efficiency increasing from 0% to $\sim 1\%$.

Breakoff Speed: Breakoff speed is slower to account for high viscosity of enzymes.

Delay: Dispense delay is increased by 10x to provide adequate time for the viscous enzyme to be dispensed completely.

7.3.0.4 DNA Liquid Class: Aspirate Protocol Parameters

Liquid Density: 1 mg/ μL

Table 7.4: Aspirate Parameters for DNA liquid class by volume-based sub-class

PARAMETER	LIQUID SUB-CLASS		
	1.5 - 3.01 μL	3.01 - 5.01 μL	5.01 - 15.01 μL
Speed	2 $\mu\text{L}/\text{s}$	5 $\mu\text{L}/\text{s}$	50 $\mu\text{L}/\text{s}$
Delay	500 ms	500 ms	1000 ms
System Trailing Airgap (STAG)	20 μL	20 μL	20 μL
Leading Airgap (LAG)	10 μL	10 μL	15 μL
Trailing Airgap (TAG)	0 μL	0 μL	1 μL
Excess Volume	0 μL	0 μL	0 μL
Conditioning Volume	0 μL	0 μL	0 μL
Pinch Valve	Off	Off	Off
Liquid Level Detection	Yes	Yes	Off
Mix before Aspiration	No	No	No
Aspirate Position	Liquid level +1 mm offset (with tracking)	Liquid level +1 mm offset (with tracking)	z-max with -1mm offset (no tracking)
Retract Tips to	current position - 5 mm	current position - 5 mm	current position - 5 mm
Retraction Speed	20 mm/s	20 mm/s	20 mm/s

Table 7.5: Dispense Parameters for DNA liquid class by volume-based sub-class

PARAMETER	LIQUID SUB-CLASS		
	1.5 - 3.01 μL	3.01 - 5.01 μL	5.01 - 15.01 μL
Speed	5 $\mu\text{L/s}$	10 $\mu\text{L/s}$	10 $\mu\text{L/s}$
Breakoff speed	10 $\mu\text{L/s}$	10 $\mu\text{L/s}$	30 $\mu\text{L/s}$
Delay	200 ms	200 ms	1000 ms
Trailing airgap after each dispense	No	No	No
Pinch Valve	No	No	No
Liquid Level Detection	Yes	Yes	No
Dispense Position	z-max with tracking (no offset)	z-max with tracking (no offset)	z-max -2 mm offset (no tracking)
Tip Touching	No	No	No
Mix after dispense	No	No	No
Retract Tips to	Current position	Current position	z-dispense
Retract Speed	50 mm/s	50 mm/s	50 mm/s

7.3.0.5 DNA Liquid Class: Dispense Protocol Parameters

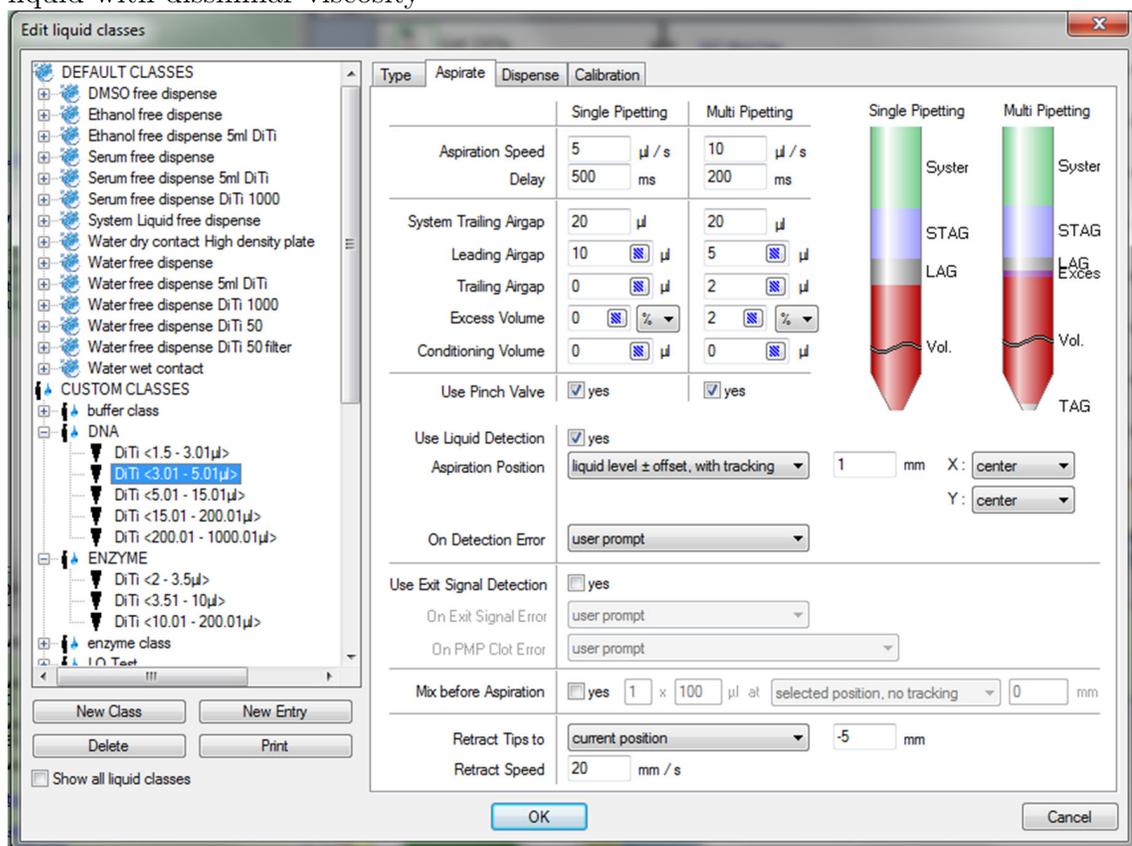
7.3.0.6 Adjustable parameters in Tecan Evo Liquid-handling robot Liquid Class Aspirate protocol:

The following are the available parameters of movement and air-displacement of the disposable pipette tip that can be customized executing aspirate parameters in an air-displacement Tecan liquid-handling robot. Figure 7.2 provides a screenshot of the aspirate parameter setup screen within the Freedom EVOware software.

7.4 Experimental validation of optimized liquid classes

The Tecan EVO 150 comes with a single predefined liquid class for pipetting water. Successful automated DNA assembly requires automated pipetting of water, DNA enzyme, and enzyme buffer with adequate accuracy and precision allowing correct clones to be generated with consistency over multiple trials. Since DNA assembly reactions in synthetic biology are carried out in small volumes, we chose to validate the liquid sub-classes for the lowest volume range functionally, by performing MoClo DNA assembly reactions. With the limitation that the Tecan EVO 150 cannot pipette volumes smaller than 2 μL with either precision or accuracy, and being

Figure 7.2: Screenshot of Aspirate liquid class optimization panel in the Tecan EVOware software. Parameters must be adjusted for very-small ($1.5 - 3.0\mu\text{L}$), small ($5.0 - 15.0\mu\text{L}$), mid-range ($15.0 - 200\mu\text{L}$) and large ($>200\mu\text{L}$) volume ranges for all liquid with dissimilar viscosity



sensitive to the cost of reagents, we set up $20\mu\text{L}$ MoClo reactions such that the minimum volume of each reagent being automatically pipetted is $2\mu\text{L}$. Additionally, we wanted to define the minimum precision and accuracy of automated liquid-handling that provided DNA assembly efficiency comparable to manual DNA assembly.

7.4.1 Automated DNA Assembly

Using the MoClo DNA assembly instructions generated using Tecan EVOware software 7.4.3, 16 MoClo DNA assembly reactions from 16 promoter, 3 RBS, 4 gene, 4

terminator, and 4 destination vector parts using reagents and reaction setup parameters defined in Table 7.6 and shown in Figure 7.3.

We collected the data for the number of positive and true positive colonies from each of the 16 circuits. 100 assumed positive (white) colonies (or the total number of available white colonies if there were less than 100 total white colonies) from the blue-white screening post automated DNA assembly were sequenced to determine if they were true positives. With the exception of one circuit (circuit 6), where one white colony was found to have no insert, all sequenced white colonies were found to be correct. The data can be found in Table 7.7.

7.4.2 Modular Cloning (MoClo) Reaction Parameters and reaction protocol:

Each of 4 DNA parts (promoter, RBS, gene and terminator) along with destination vector (DV) were diluted to 10 fmol concentrations and added sequentially, in equimolar, amounts to the reaction. DNA parts can be purified plasmid DNA, PCR fragment or synthesized double-stranded DNA. DNA parts were combined with 1x T4 DNA ligase buffer (Promega), 20 units of the Type IIS restriction endonuclease, BsaI (New England Biolabs) and 800 units of T4 DNA ligase enzyme (New England BioLabs). The reaction was brought up to a final volume of 20 μ L. The MoClo reaction setup is shown below in Table 7.6. MoClo reactions were performed in an Eppendorf Mastercycler ep thermocycler (Eppendorf) using the following parameters: 25 cycles (37°C for 1.5 minutes, 16°C for 3 minutes), 50°C for 5 minutes and 80° for 10 minutes. Reactions were then held at 4° until transformed.

Table 7.6: **Modular Cloning (MoClo) reaction setup.** MoClo DNA assembly component volumes for a single 20 μ L MoClo reaction.

Reaction per well	
Deionized Water	4 μ L
10x Ligase Buffer	2 μ L
Promoter (10 fmol)	2 μ L
RBS (10 fmol)	2 μ L
Gene (10 fmol)	2 μ L
Terminator (10 fmol)	2 μ L
Destination Vector (10 fmol)	2 μ L
BsaI (20 U)	2 μ L
T4 DNA Ligase (800 U)	2 μ L
TOTAL	20 μ L

7.4.3 Automated DNA assembly protocol and setup

Using reagent and DNA part locations on Reagents and DNA source plates respectively, a series of instructions for carrying out the MoClo reactions were generated using the Tecan EVOware drag-and-drop software. Each MoClo reaction was set up as described in Table 7.6. The Tecan EVOware instructions file, Automated_MoClo.esc for the assembly is publicly available [18].

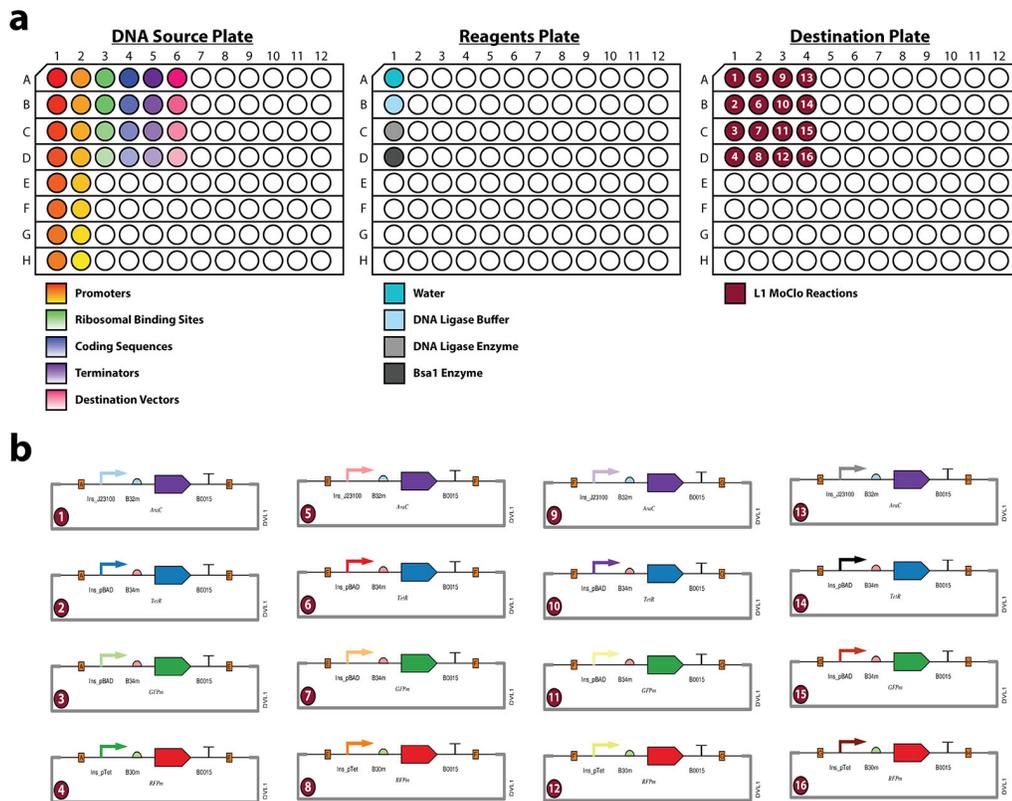
Reagents (water, buffer and enzyme) and diluted DNA parts were setup in separate skirted 96-well PCR plates (See Figure 7.3). The reagent plate was further placed in an 96-well, iceless, PCR storage cold block (Eppendorf). A third skirted 96-well PCR plate was used as the destination plate for the final MoClo reactions. All labware (including the Eppendorf cold block) and labware locations on the Tecan EVO 150 bed were pre-defined in the EVOware protocol prior to executing the DNA assembly protocol. Locations for waste was also pre-defined prior to executing DNA assembly. Reagents and DNA parts were aspirated and dispensed into the destination plate

in the following order: water, buffer, RBS, gene, terminator, destination vector, promoter, DNA ligase, and BsaI enzyme.

Table 7.7: Positive and True Positive Colony count from automated DNA assembly by the MoClo assembly method on the Tecan EVO 150 using optimized liquid classes using automated MoClo reaction setup outlined in Table 7.6.

	# total colonies	# white colonies (assumed positive)	# true positive/100 white colonies
Circuit 1	321	309	100
Circuit 2	290	273	100
Circuit 3	187	178	100
Circuit 4	255	247	100
Circuit 5	417	414	100
Circuit 6	124	77	100
Circuit 7	332	320	99
Circuit 8	210	203	100
Circuit 9	255	249	100
Circuit 10	343	329	100
Circuit 11	198	188	100
Circuit 12	265	249	100
Circuit 13	283	280	100
Circuit 14	184	176	100
Circuit 15	201	189	100
Circuit 16	192	189	100

Figure 7.3: **Setup for Reagent, DNA parts and MoClo reaction plates in automated DNA assembly.** Individual DNA parts from the DNA plate are used to construct fully assembled transcriptional units in the Destination Plate. (a) Plate setup: DNA Source Plate contains promoters (columns 1-2), RBSs (column 3), CDSs (column 4), terminators (column 5) and destination vectors (column 6). Reagent Plate contains water, 10x ligase buffer, DNA ligase enzyme and Bsa I enzyme in rows A-D in column 1 respectively. Destination plate shows the final location for each unique MoClo reaction. The number in each well in the Destination Plate corresponds to the Level 1 MoClo circuit in Figure 7.3(b). (b) 16 unique Level 1 MoClo circuits built automatically.



Chapter 8

Conclusions and Future Work

Synthetic biology is premised on the principle that genetic parts are modular and composable. However, the re-composition of genetic parts into devices alters the DNA sequence at the junctions of genetic parts, thus altering the genetic context of these parts. Recently the wide variability in expression of promoters [38] and ribosome binding sites (RBSs) [50] due to fluctuating genetic context has been under scrutiny. For example, Bicistronic design (BCD) parts that regularize fluctuations at the promoter:RBS junction by adding a second RBS element translating a short peptide sequence upstream of the RBS recruiting ribosomes to the gene of interest are starting to replace traditional RBS parts. These BCD parts are demonstrably more predictable in their behavior. To further eliminate unexpected circuit behavior as a result of genetic context effects, we must also consider the general design of promoter parts used to create synthetic genetic circuits.

8.1 Replacing minimal constitutive promoters with consistent constitutive promoters

This work along with a previous study [38] showed that a number of the minimal constitutive bacterial promoter parts used in synthetic biology show context dependence when just 1-4 base pairs upstream of the promoter 5' boundary is altered.

This context dependence can be eliminated by using screened insulator sequences as described in this study (see Figure 5.1, Figure 5.2, Figure 5.7 and, Figure 5.8).

One possible method to avoid promoter genetic context effects in a target engineered genetic circuit would be to evaluate each promoter within the target circuit and evaluate whether the promoter is likely to be susceptible to genetic context fluctuations using the insulation workflows provided here (See Figure 8.1 and Figure 8.3). This approach has higher cost (see Figure 5.1 and Figure 5.2) in terms of both time and resources: an insulator screen requires expensive, non-reusable primers and will add a minimum of 8 days to the build-cycle. A further disadvantage would be that if insulation is added to a promoter within its final context, no new insulated promoter parts are generated. If the same promoter needs to be insulated for a different device, a new promoter screen must be initiated.

A better approach is to convert all commonly used minimal promoters and promoters with RNA polymerase or transcription factor binding sites immediately downstream of the 5' promoter boundaries into reusable promoter parts by screening each promoter once and then using the screen to determine the DNA sequence of the spacer/insulator that is incorporated into the promoter. This requires a one-time insulated promoter library screening investment and generates reusable promoter parts. To this end, all promoters used in this study are being submitted to Addgene [1] and the Registry of Standard Biological Parts [35] to share with the larger synthetic biology community.

Workflows for predicting the need for insulation are not perfect: a preliminary investigation of the pBAD promoter architecture would lead one to choose to not insulate promoter pBAD if there was only a single instance of the promoter in a design (See Figure 8.1. Yet, pBAD is significantly affected by upstream genetic context (see

Figure 4.8) that is alleviated by insulation.

When the J231XX minimal constitutive promoter series was created, promoter genetic context effects had not been documented, or even anticipated. Before conclusive evidence on the pervasiveness of genetic context effects caused by fluctuating DNA sequence at part junctions was available, the standard format for creating synthetic genetic parts had already been defined as only the stretch of DNA base pairs that included all necessary features for part function only. DNA bases to insulate parts against genetic context were not considered a necessary component in genetic part design. Given the frequency with which promoter genetic context effects are seen when standard promoter parts are recomposed into engineered genetic circuits, the current design of promoter parts for bioengineering may be “too minimal”. As in the case of RBS to BCD design, or RBS to ribozyme+RBS design as better alternatives to RBS parts, it would be beneficial to amend promoter part design protocols to include a screened insulator/spacer element as part of general part creation procedure.

8.1.1 Promoter Insulation Workflow

Insulating promoters adds time and supply costs to the process of creating genetic circuits. Therefore, the decision on whether or not a promoter in a given design should be insulated or left as is must be carefully evaluated. The decision tree for evaluating whether or not a given promoter should be insulated is presented in Figure 8.1. The most widely used promoter parts in synthetic biology were designed in a way that did not include any additional nucleotides adjacent to the promoters’ protein binding sites - whether RNA Polymerase (RNAP) binding sites or transcription factor binding sites. When a promoter is fused to a new part, the DNA sequence immediately upstream (5’) of the protein binding sites changes and likely impacts RNA

polymerase enzyme (in the case of minimal constitutive promoters) or transcription factor (in the case of regulated promoters) binding in some way that alters promoter expression. A detailed investigation of the mechanism by which these effects occur would be illuminating but was beyond the scope of this investigation.

The minimal requirements for executing this decision tree are that there must be an unambiguous genetic circuit design, and that all promoter parts within this design have been assigned physical (promoter) part samples with known DNA sequences for which basic characterization data is available. Each promoter must be evaluated separately. Once the decision to insulate a promoter part has been made, the specific steps for insulation follow the workflow outlined in Figure 8.2. The process of primer design can be automated using a PERL script (See Appendix)

It has been shown that many (but not all) minimal promoters (-35 RNAP binding site, -10 RNAP binding site and intervening sequences between the -35 and -10 sequences) behave unpredictably when the DNA sequence immediately 5' of the promoter boundary is altered [38]. Minimal promoters can be left uninsulated in one of three cases:

1. If the 4 base pair DNA sequence immediately upstream of the minimal promoter part in the final device is identical to that of the context in which the promoter was characterized.
2. If the minimal promoter has been characterized in its final context of DNA sequence at its 5 boundary and found to behave identically to the version of itself in its original characterization context.
3. If characterization data for the precise promoter and upstream DNA sequence in which the promoter is to be recomposed as part of the final circuit design is

available and is being used in modeling the final device behavior.

However, as has been mentioned above, most minimal promoters are sensitive to DNA sequence variations immediately upstream of their 5' boundaries. Therefore, in the absence of detailed characterization data for the minimal promoter in question with the precise final context DNA sequence, it will likely save time to simply insulate minimal promoters as a routine course of action. Thus the insulation workflow for this step can be simplified by eliminating this decision step (blue diamond) entirely. For promoters in a design that are not defined as minimal promoters, it must first be determined based on the promoter architecture whether the 5' boundary of the promoter part is directly adjacent to a protein binding site. (Note: minimal promoters are a subset of this type of promoter where there is always a protein binding site immediately adjacent to the promoter 3' boundary.) Variations in the DNA sequence immediately upstream of protein binding sites can alter the kinetics of transcription factor binding as explained above. Promoters in which transcription factor binding sites are positioned adjacent to the 5' promoter boundary are likely to be impacted by fluctuations in genetic context and should be insulated. However, as in the case of the repressible pTet promoter, some promoters that fall within this category may not be impacted by changes in DNA sequence immediately upstream of the promoter transcription factor binding sites for the specific upstream sequences in a particular final design. This was the case in my 24 inverter designs, where the pTet promoter was verified to behave to behave identically regardless of whether MoClo scar A (5-GGAG-3), E (5GCTT-3), F (5-CGCT-3) or G (5-TGCC-3) scar was placed upstream of it. This was determined by characterizing the pTet promoter in isolation with each upstream scar. As genetic circuits become larger and more complex, insulating promoters that have protein binding sites adjacent to the 5 promoter boundary

may be a more practical and less time consuming step rather than testing out all promoters with each specific upstream DNA sequence before deciding to insulate the promoter. Again, the insulation workflow can be simplified by eliminating this decision point and simply insulating a promoter that is used multiple times, and with varying upstream DNA sequences, within a design.

Similarly, if multiple instances of a promoter are being used within a single genetic device, insulating the promoter is likely the only way to guarantee that each instance behaves identically regardless of where in the device it is placed without resorting to characterizing the promoter with each upstream DNA sequence.

As such, Figure 8.3 represents the simplified workflow to be followed in deciding whether to insulate promoters in a given design without having additional characterization steps for promoters from the design in isolation. The simplified workflow is more in keeping with the the principles of composability that is central to the philosophy of synthetic biology. It also eliminates the need for collecting characterization data in nonstandardized formats (vector backbones, upstream DNA sequences, antibiotic cassettes).

8.1.1.1 Insulation workflow applied to promoters from 24 inverter permutations

Here I demonstrate the workflow as applied to the three promoters (J23100, pBAD and pTet) from my 24 inverter permutations. Steps that were not necessary because of the decisions taken according to the particular promoter are shown in gray. Figure 8.4 depicts the promoter insulation workflow applied to the minimal constitutive promoter J23100. Had I not been testing all 24 permutations, I would not have needed to insulate promoter J23100 as the genetic context (4 bp sequence upstream

of the 5' promoter boundary) is the same as that in which the promoter was originally characterized. However, as I was also making inverters where the J23100 was positioned elsewhere in the device, I need to insulate it.

Figure 8.5 shows the application of the insulation workflow to promoter pBAD. Promoter pBAD is used twice in the device with varying genetic context between the two instances of the promoter. Thus, even in the absence of characterization data for pBAD_A, pBAD_-, pBAD_F and pBAD_G, the insulation provides the safest route to getting equal expression from both instances.

Promoter pTet is an example where the workflow guides us towards insulating a promoter where it was not strictly necessary. According to the workflow figure for promoter pTet (Figure 8.6), pTet should be insulated. However, according to the characterization data for promoter pTet in isolation (pTet_A-RFP, pTet_E-RFP, pTet_F-RFP, pTet_G-RFP), we know that there is no significant difference in expression when the 4 base pairs upstream of promoter pTet is varied. Insulating pTet improves performance (in addition to insulating promoters J23100 and pBAD, Figure 5.8) over insulating just promoters J23100 and pBAD (Figure 6.2.3) but is not essential to eliminating order dependence.

8.1.2 Applying DNA spacer screening in other organisms

Fluctuating promoter expression as a result of changing DNA sequence at promoter 5' junctions has been extensively studied in *E. coli*. However, the problem has been known to exist in other organisms as well (personal communication, Karmella Haynes, Arizona State University, Synthetic Biology Engineering Research Center Spring Retreat, 2015). Synthetic promoter parts in other host organisms may also benefit from insulation against their upstream sequences. Procedurally, there are no

barriers to transferring the DNA spacer generation and screening methodology to unidirectional promoter parts in other species of bacteria, as well as in yeast and even mammalian systems. No modifications to the methodology need to be made. Insulated promoter libraries can be grown for screening using the standard growth protocols for flow cytometry for that species and type of promoter. DNA spacers can be inserted upstream of eukaryotic minimal and other constitutive promoters in exactly the same way as in the bacterial constitutive promoter (See Figure 5.1). In mammalian regulated promoters the DNA spacer methodology could serve the additional purpose of adjusting promoter expression levels without having to change the locus of cloned regulator elements (such as enhancer or repressor elements), thus reducing the number of design, build, test cycles. However, the spacer screening methodology cannot be applied to bidirectional promoters [55, 69]. Bidirectional promoters are short (<1kb) regions of intergenic DNA between the 5' ends of two divergent genes on either strand of DNA. The intergenic region serves as a promoter element for both genes and adding a spacer of either side of the promoter would disrupt transcription initiation from the gene on the opposite DNA strand.

8.1.3 Mining DNA spacers for an improved DNA spacer development process

A detailed analysis of the sequence and structural properties of DNA spacer sequences was beyond the scope of the project presented in this thesis. Given the immense size (4^{36}) of the theoretical design space of 36 base pair DNA spacers along with the wide distribution of expression levels observed in each even the small, 200-sample spacer screens, it appears likely that the DNA spacer sequences obtained through screening are information-rich and can be mined for connections between specific

DNA sequence motifs or secondary structures to a spacer's impact on promoter expression. This, in turn, could result in design rules that would allow synthetic biologists to eventually rationally design DNA spacers rather than having to screen a library of candidates.

In the absence of a canonical bacterial DNA spacer sequence, any investigation into sorting screened DNA spacers would necessitate heuristic analyses of much larger number of candidates. Candidate samples from a DNA spacer library can be sorted during flow cytometry screening using a cell sorter and divided into a predefined number of clusters based on expression level relative to promoter expression in its original characterization context. The entirety of screened candidate spacers would have to be sequenced, which would likely require deep (Next Generation) sequencing methods. Samples within each cluster can then be analyzed for common motifs, secondary structures, DNA looping etc. Conserved commonalities within each cluster can be formalized into rules for spacer design. These rules would then need to be tested by testing the ability of rationally designed DNA spacers created using these rules to precisely and reliably insulate promoters. Once verified, these rules can be incorporated into bio-design software programs such as EUGENE as a standard step in the design process for genetic circuits.

Figure 8.1: **Decision process workflow for promoter insulation.** The flowchart depicts the critical decision points in evaluating whether or not a promoter part to be included in a final design needs to be insulated. A promoter part of known DNA sequence and position within final device design is the input for the workflow. The workflow must be executed in its entirety for each promoter part within a genetic circuit design. Blue decision boxes represent steps that are removed in the simplified workflow. Green curved rectangles represent terminal decision. Terminal decision is distinct from output which is either an insulated promoter or the original promoter left unaltered.

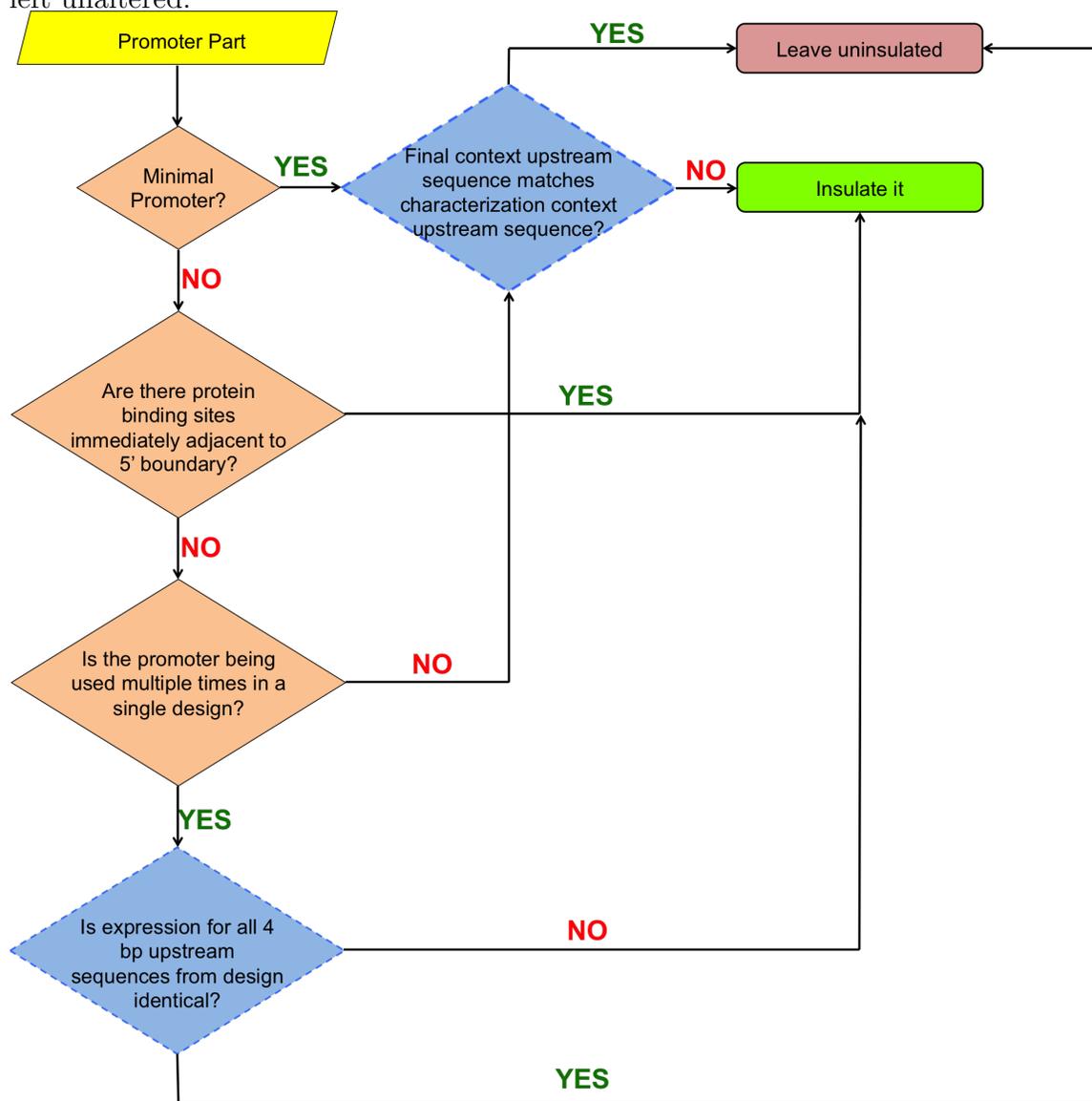


Figure 8.2: **High-level schematic for promoter insulation process.** The flowchart outlines the process for promoter insulation described in Figure 3.3 (Chapter 3). The input (yellow parallelogram) is an uninsulated promoter part. The output (green parallelogram) is an insulated promoter part. Curved-edge rectangles represent steps in the protocol. The primer design step can be automated (green curved-edge rectangle). Purple curved-edge rectangle steps must be performed manually. DNA assembly using the insulated promoter part obtained as the output can be automated using a liquid-handling robot using Modular Cloning (MoClo) using protocols described in Chapter 7. PERL script to generate primers is provided in Chapter 10.

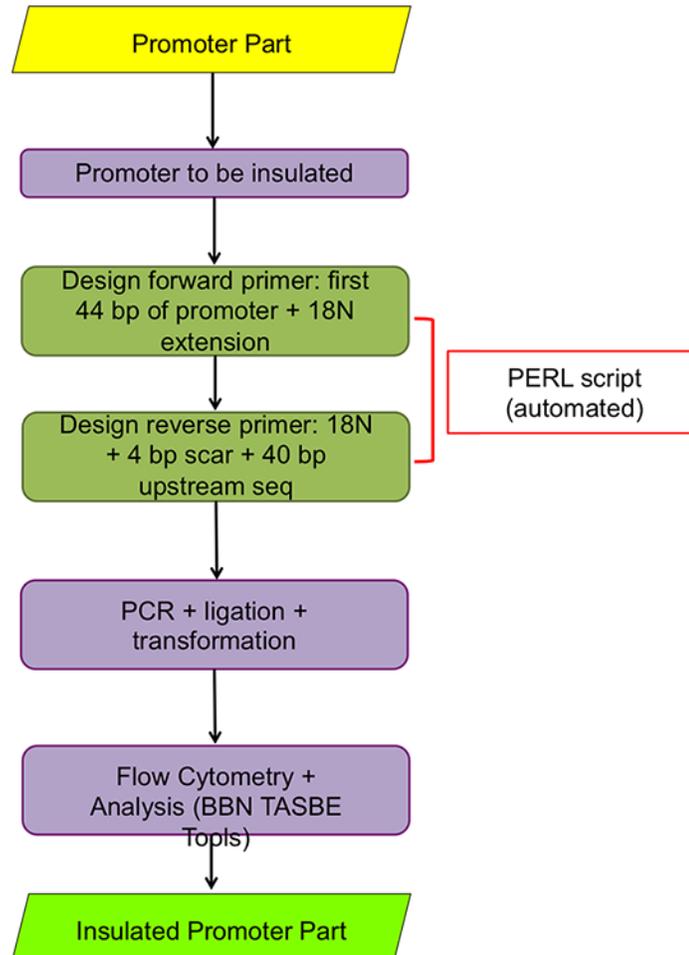


Figure 8.3: **Simplified decision process workflow for promoter insulation.** The flowchart depicts the decision points in evaluating whether or not a promoter part to be included in a final design needs to be insulated. A promoter part of known DNA sequence and position within final device design is the input for the workflow. The workflow must be executed in its entirety for each promoter part within a genetic circuit design. Blue decision boxes from Figure 8.1 have been removed. Terminal decision is distinct from output which is either an insulated promomoter or the original promoter left unaltered.

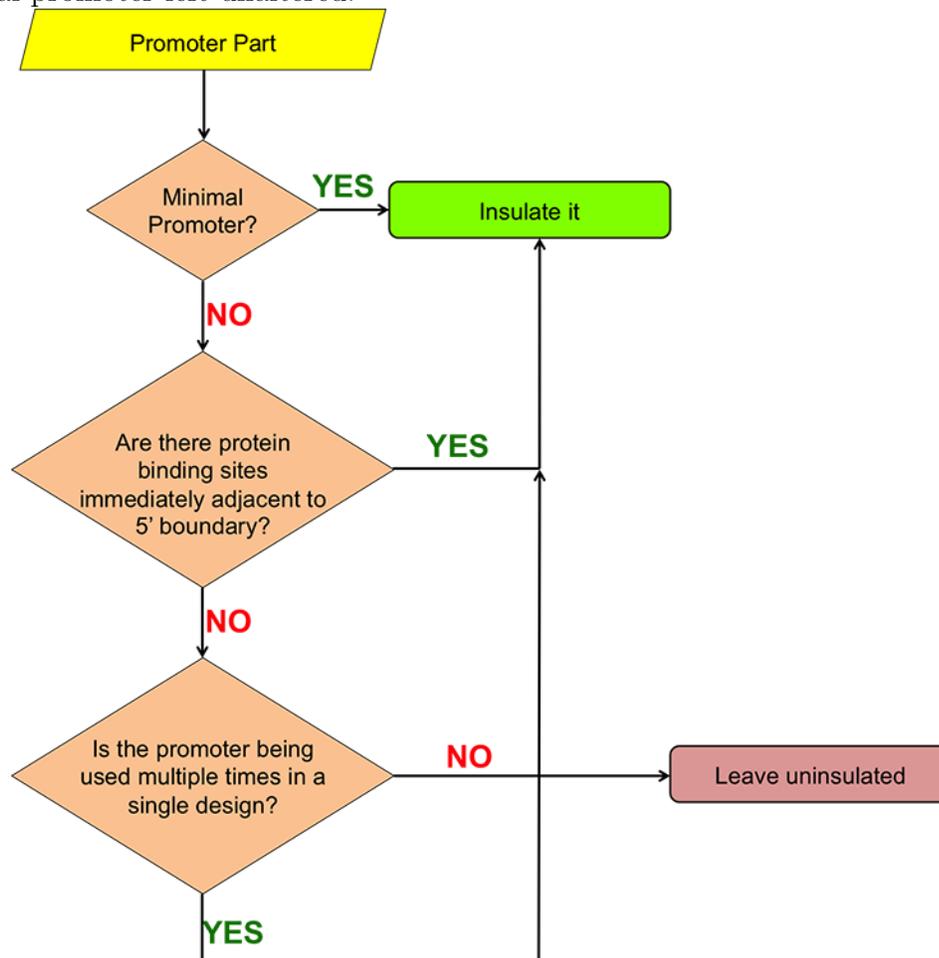


Figure 8.4: **Promoter insulation protocol applied to constitutive minimal promoter J23100.** The flowchart depicts the decision points that were evaluated in insulating promoter J23100 from 24 inverter circuits. All decision points that were not evaluated are shown in gray. Based on both workflow and characterization data for promoter J23100 (Figure 4.7), I decided to insulate it.

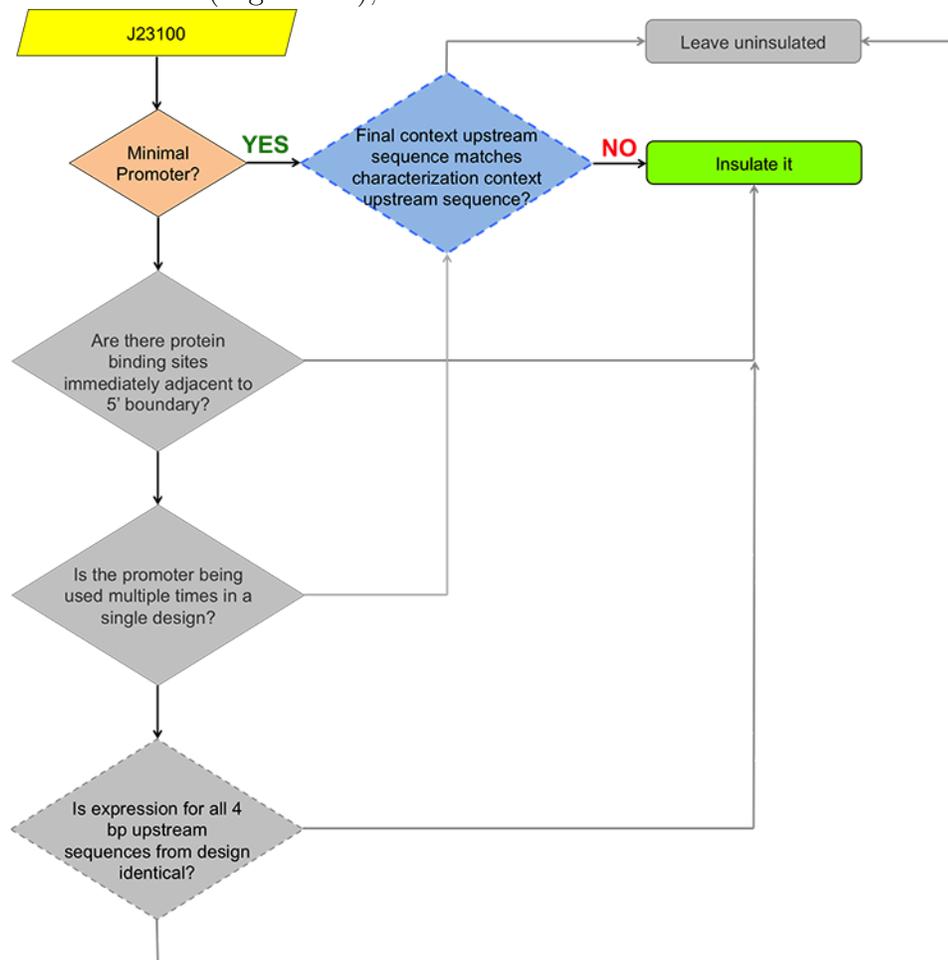


Figure 8.5: **Promoter insulation protocol applied to inducible promoter pBAD.** The flowchart depicts the decision points that were evaluated in insulating promoter pBAD from 24 inverter circuits. All decision points that were unnecessary are shown in gray. Based on both workflow and characterization data for promoter pBAD (Figure 4.8), I decided to insulate it.

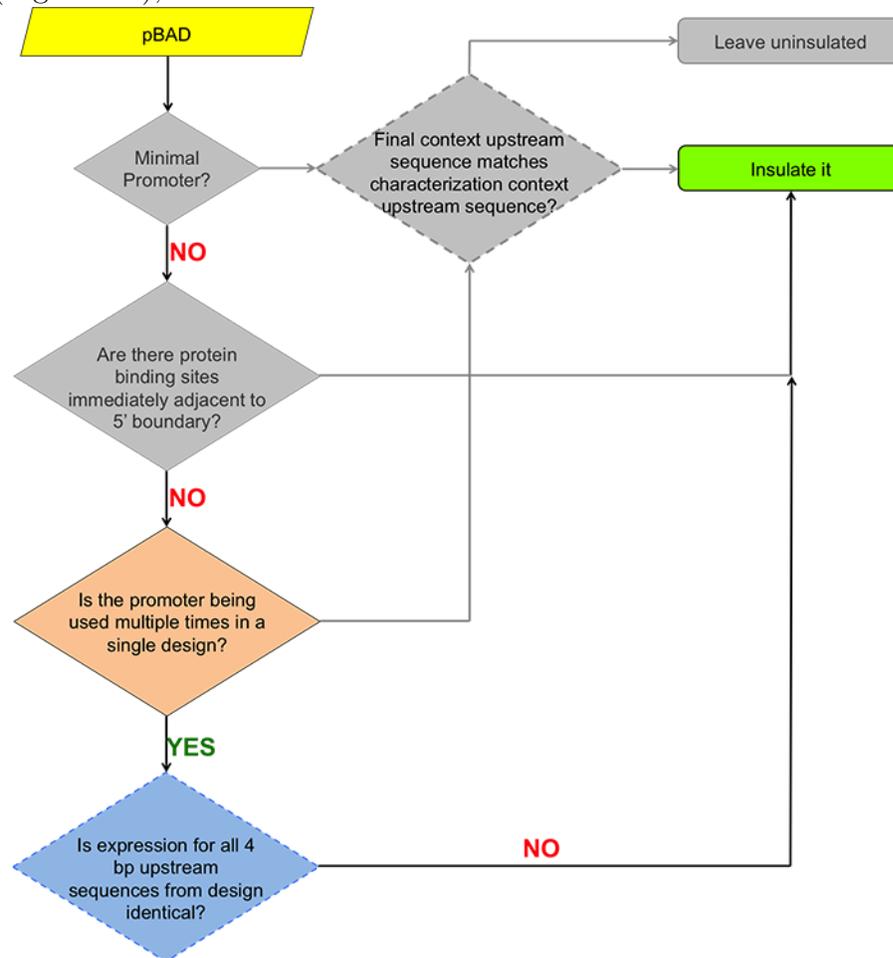
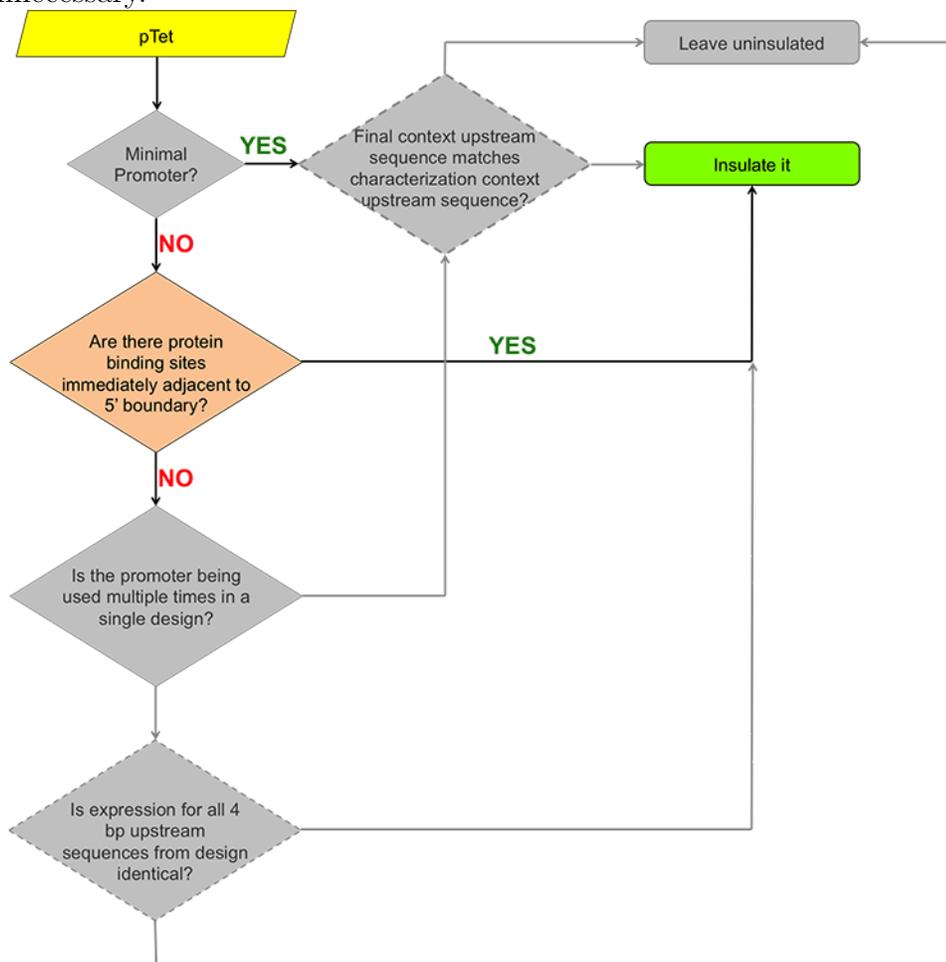


Figure 8.6: **Promoter insulation protocol applied to repressible promoter.** The flowchart depicts the decision points that were evaluated in insulating promoter pTet from 24 inverter circuits. All decision points that were unnecessary are shown in gray. Based on characterization data for promoter pTet (Figure 4.9), insulating it was unnecessary.



Chapter 9

Appendix A: 24 Inverter Permutations

Raw Data

Table 9.1: **GFP and RFP expression of uninsulated 24 inverter permutations upon induction with 1mM L-Arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for uninsulated 24 inverter set upon induction with L-arabinose for 14 hours. Data graphed in Figure 4.2.

ARABINOSE +				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std.Dev. of Mean (fold)
Inverter-01	668.89	1.02	160.70	1.10
Inverter-02	621.31	1.01	145.68	1.00
Inverter-03	656.37	1.01	155.71	1.03
Inverter-04	598.06	1.00	143.93	1.00
Inverter-05	2096.02	1.30	14220.90	1.65
Inverter-06	667.62	1.04	298.63	1.14
Inverter-07	728.24	1.01	12803.54	1.65
Inverter-08	735.93	1.00	7683.69	3.01
Inverter-09	1575.55	1.24	1313.00	1.37
Inverter-10	1418.81	2.01	512.99	3.30
Inverter-11	42384.40	1.07	202.64	1.02
Inverter-12	806.84	1.04	9655.75	1.48
Inverter-13	3375.55	1.10	2895.00	1.36
Inverter-14	16814.54	1.40	4131.88	1.21
Inverter-15	10740.46	2.76	2692.28	10.90
Inverter-16	6576.17	3.75	159.16	1.07
Inverter-17	16880.06	1.06	187.97	1.03
Inverter-18	14812.93	1.09	40315.50	1.13
Inverter-19	884.28	1.01	37982.10	1.04
Inverter-20	630.15	1.02	155.31	1.05
Inverter-21	860.93	1.03	42550.53	1.09
Inverter-22	11869.87	1.30	32382.55	1.44
Inverter-23	848.01	1.02	41914.86	1.10
Inverter-24	848.29	1.01	40183.89	1.11

Table 9.2: **GFP and RFP expression of uninsulated 24 inverter permutations in the absence of induction with 1mM L-Arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for uninsulated 24 inverter set in the absence of induction with L-arabinose for 14 hours. Data graphed in Figure 4.2.

ARABINOSE -				
Device ID	GFP		RFP	
	Mean	Std.Dev. of mean (fold)	Mean	Std.Dev. of mean (fold)
Inverter-01	647.35	1.04	244.17	1.15
Inverter-02	640.91	1.04	156.42	1.08
Inverter-03	683.12	1.14	338.03	2.06
Inverter-04	598.18	1.02	148.42	1.02
Inverter-05	717.68	1.04	16168.21	1.02
Inverter-06	618.15	1.03	534.18	1.16
Inverter-07	599.84	1.01	1396.71	2.72
Inverter-08	587.09	1.02	3528.13	5.53
Inverter-09	641.22	1.05	8158.91	2.41
Inverter-10	699.43	1.22	3456.13	14.60
Inverter-11	4679.83	1.92	267.00	1.31
Inverter-12	682.37	1.10	4577.52	2.92
Inverter-13	741.73	1.09	11264.40	1.34
Inverter-14	746.51	1.04	10596.61	1.81
Inverter-15	1279.69	1.20	4270.28	8.33
Inverter-16	1001.59	1.64	174.58	1.12
Inverter-17	806.32	1.07	174.76	1.03
Inverter-18	2054.68	1.24	47835.69	1.72
Inverter-19	1118.74	1.09	34077.76	1.33
Inverter-20	639.35	1.02	154.28	1.07
Inverter-21	767.88	1.08	39973.21	1.36
Inverter-22	3079.63	1.72	24336.33	1.72
Inverter-23	664.18	1.05	26338.01	1.42
Inverter-24	649.60	1.03	31289.67	1.34

Table 9.3: **GFP and RFP expression of 24 inverter permutations with all rationally insulated promoters upon induction with 1mM L-Arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for rationally insulated 24 inverter set upon induction with L-arabinose for 14 hours. Data graphed in Figure 4.13.

ARABINOSE +				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	8764.36	1.25	3316.94	1.40
Inverter-02	6441.78	1.53	2850.53	1.43
Inverter-03	21967.72	1.23	13391.99	1.36
Inverter-04	634.46	1.02	154.03	1.06
Inverter-05	661.84	1.05	147.06	1.01
Inverter-06	721.82	1.05	153.90	1.10
Inverter-07	18375.03	1.18	5495.46	1.35
Inverter-08	19220.16	1.12	12286.44	1.67
Inverter-09	15458.56	1.12	4179.09	1.06
Inverter-10	26292.45	1.13	19629.13	1.01
Inverter-11	21333.94	1.11	3044.67	1.25
Inverter-12	688.46	1.03	174.56	1.03
Inverter-13	3423.62	6.06	1654.74	6.25
Inverter-14	1031.36	5.61	341.85	8.89
Inverter-15	21662.87	1.10	17513.74	1.35
Inverter-16	20280.70	1.23	22495.05	2.04
Inverter-17	3296.64	5.55	5736.26	13.06
Inverter-18	1246.94	1.34	152.88	1.01
Inverter-19	3049.71	9.61	216.13	1.56
Inverter-20	3065.84	1.90	235.97	1.21
Inverter-21	30646.16	1.24	2803.21	1.06
Inverter-22	19274.85	1.05	3323.62	1.11
Inverter-23	1371.63	2.51	204.07	1.28
Inverter-24	1556.73	2.33	187.67	1.17

Table 9.4: **GFP and RFP expression of 24 inverter permutations with all rationally insulated promoters in the absence of induction with L-arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for rationally insulated 24 inverter set in the absence of induction with L-arabinose for 14 hours. Data graphed in Figure 4.13.

ARABINOSE -				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	18819.41	1.30	6655.84	1.38
Inverter-02	2374.86	1.16	8730.66	1.35
Inverter-03	4162.54	1.04	17512.53	1.11
Inverter-04	684.05	1.09	204.79	1.35
Inverter-05	655.06	1.01	154.80	1.03
Inverter-06	664.60	1.05	157.22	1.08
Inverter-07	40602.33	1.07	13149.61	1.14
Inverter-08	3582.53	1.10	19243.72	1.00
Inverter-09	1225.67	1.22	5330.39	1.41
Inverter-10	1688.70	1.04	15660.04	1.10
Inverter-11	2409.97	1.07	7590.28	1.18
Inverter-12	694.62	1.06	194.25	1.16
Inverter-13	1540.33	2.55	2128.25	8.61
Inverter-14	1076.39	1.74	588.72	6.18
Inverter-15	1627.08	1.02	16283.62	1.09
Inverter-16	1308.17	1.16	15240.70	1.38
Inverter-17	1551.53	2.37	4159.03	14.17
Inverter-18	878.91	1.07	166.69	1.10
Inverter-19	2610.13	4.64	240.39	1.51
Inverter-20	1499.63	1.68	312.35	1.08
Inverter-21	35827.71	1.06	6051.78	1.24
Inverter-22	2056.12	1.04	9616.70	1.11
Inverter-23	1042.61	1.57	237.73	1.35
Inverter-24	1310.98	1.70	249.83	1.40

Table 9.5: **GFP and RFP expression of promoter J2300-only insulated 24 inverter permutations upon induction with 1mM L-Arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for promoter J23100-only insulated 24 inverter set in the presence of induction with L-arabinose for 14 hours. Data graphed in Figure 6.1.

ARABINOSE +				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	549.54	1.08	6918.31	2.73
Inverter-02	20417.38	2.98	933.25	1.58
Inverter-03	10232.93	1.44	1318.26	1.13
Inverter-04	562.34	1.05	169.82	1.05
Inverter-05	537.03	1.01	165.96	1.02
Inverter-06	562.34	1.02	18620.87	1.03
Inverter-07	537.03	1.02	7585.78	1.30
Inverter-08	22908.68	1.64	1122.02	1.24
Inverter-09	2630.27	1.34	1513.56	1.35
Inverter-10	17782.79	1.07	1737.80	1.29
Inverter-11	30199.52	2.59	977.24	1.30
Inverter-12	575.44	1.04	22387.21	1.14
Inverter-13	13489.63	1.46	912.01	1.12
Inverter-14	13803.84	1.83	1202.26	1.41
Inverter-15	1548.82	1.19	467.74	1.21
Inverter-16	7762.47	2.66	794.33	2.10
Inverter-17	23988.33	1.04	2137.96	1.06
Inverter-18	20892.96	1.05	1659.59	1.25
Inverter-19	602.56	1.05	17378.01	1.06
Inverter-20	67608.30	1.02	3467.37	1.03
Inverter-21	549.54	1.10	19498.45	1.11
Inverter-22	95499.26	1.02	3548.13	1.05
Inverter-23	41686.94	1.06	2818.38	1.03
Inverter-24	38904.51	1.13	2454.71	1.04

Table 9.6: **GFP and RFP expression of promoter J23100-only insulated 24 inverter permutations in the absence of induction with 1mM L-arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for promoter J23100-only insulated 24 inverter set in the absence of induction with L-arabinose for 14 hours. Data graphed in Figure 6.1.

ARABINOSE -				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	562.34	1.02	9332.54	4.20
Inverter-02	7079.46	1.13	11220.18	1.15
Inverter-03	1202.26	1.10	19054.61	1.10
Inverter-04	549.54	1.04	218.78	1.27
Inverter-05	537.03	1.02	165.96	1.06
Inverter-06	588.84	1.03	18197.01	1.16
Inverter-07	549.54	1.02	33113.11	1.05
Inverter-08	8511.38	1.20	16218.10	1.10
Inverter-09	741.31	1.09	20417.38	1.05
Inverter-10	1621.81	1.30	31622.78	1.54
Inverter-11	8128.31	1.35	22387.21	1.20
Inverter-12	562.34	1.05	39810.72	1.10
Inverter-13	2238.72	1.13	19054.61	1.08
Inverter-14	2290.87	1.12	12022.64	1.06
Inverter-15	741.31	1.07	28183.83	1.57
Inverter-16	977.24	1.31	40738.03	1.16
Inverter-17	1737.80	1.19	16218.10	1.04
Inverter-18	1698.24	1.29	18620.87	1.23
Inverter-19	588.84	1.05	24547.09	1.15
Inverter-20	3801.89	1.08	18197.01	1.13
Inverter-21	616.60	1.07	32359.37	1.22
Inverter-22	5128.61	1.22	15135.61	1.05
Inverter-23	1445.44	1.39	26302.68	1.27
Inverter-24	3311.31	1.26	14454.40	1.12

Table 9.7: **GFP and RFP expression of inducible promoter pBAD-only insulated 24 inverter permutations upon induction with 1mM L-Arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for promoter pBAD-only insulated 24 inverter set upon induction with L-arabinose for 14 hours. Data graphed in Figure 6.2.

ARABINOSE +				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	713.60	1.03	316.76	1.12
Inverter-02	1915.69	1.39	306.23	1.05
Inverter-03	968.61	1.13	336.84	1.11
Inverter-04	4017.92	1.25	389.40	1.06
Inverter-05	3809.23	1.79	602.41	1.32
Inverter-06	6538.00	1.88	754.41	1.17
Inverter-07	31633.19	1.43	541.99	8.63
Inverter-08	14444.48	1.26	22037.89	1.18
Inverter-09	5456.82	1.91	9816.94	3.01
Inverter-10	1468.35	1.16	553.73	1.16
Inverter-11	9366.41	1.09	17803.32	1.31
Inverter-12	1424.60	1.12	482.65	1.03
Inverter-13	2895.20	1.20	23674.35	1.08
Inverter-14	15089.71	1.39	22825.38	1.08
Inverter-15	2369.40	1.23	36665.45	1.02
Inverter-16	1154.81	1.12	770.70	1.15
Inverter-17	17839.55	1.59	24637.46	1.15
Inverter-18	1634.86	1.07	568.47	1.00
Inverter-19	16823.18	1.34	26884.32	1.25
Inverter-20	8868.68	2.09	26416.60	1.07
Inverter-21	1072.51	1.31	433.01	1.43
Inverter-22	26318.07	1.18	18125.54	1.17
Inverter-23	970.42	1.15	369.99	1.20
Inverter-24	17312.59	1.41	22440.68	1.38

Table 9.8: **GFP and RFP expression of inducible promoter pBAD-only insulated 24 inverter permutations in the absence of induction with 1mM L-arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for promoter pBAD-only insulated 24 inverter set in the absence of induction with L-arabinose for 14 hours. Data graphed in Figure 6.2.

ARABINOSE -				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	662.68	1.03	281.47	1.08
Inverter-02	762.26	1.03	280.77	1.04
Inverter-03	707.61	1.01	364.82	1.11
Inverter-04	806.23	1.04	454.15	1.15
Inverter-05	831.96	1.11	892.54	1.65
Inverter-06	793.36	1.07	950.71	1.30
Inverter-07	2140.87	1.13	548.83	4.52
Inverter-08	1688.15	1.09	116005.61	1.08
Inverter-09	873.29	1.12	17766.75	3.94
Inverter-10	740.21	1.01	1376.56	1.17
Inverter-11	1156.42	1.00	72212.10	1.59
Inverter-12	795.29	1.08	927.97	1.09
Inverter-13	783.71	1.08	46378.53	2.08
Inverter-14	972.46	1.07	87885.66	1.21
Inverter-15	729.31	1.04	152610.00	1.03
Inverter-16	740.70	1.03	1409.01	1.51
Inverter-17	1001.38	1.01	83134.25	1.04
Inverter-18	806.18	1.04	1052.52	1.11
Inverter-19	995.67	1.05	116439.68	1.36
Inverter-20	886.97	1.04	108697.61	1.23
Inverter-21	668.56	1.01	806.05	1.50
Inverter-22	1333.06	1.10	120009.33	1.12
Inverter-23	664.71	1.04	553.10	1.39
Inverter-24	1247.93	1.04	162992.27	1.07

Table 9.9: **GFP and RFP expression of promoters J23100 and pBAD-insulated 24 inverter permutations on induction with 1mM L-arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for promoters J23100 and promoter pBAD insulated 24 inverter set upon induction with L-arabinose for 14 hours. Data graphed in Figure 6.2.3.

ARABINOSE +				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	6165.9	1.6	4941.2	1.4
Inverter-02	34735.0	1.1	3999.3	1.1
Inverter-03	8064.9	1.6	6356.7	1.3
Inverter-04	37582.4	1.1	3803.3	1.1
Inverter-05	29379.4	1.2	4083.6	1.1
Inverter-06	48585.5	1.1	4407.0	1.4
Inverter-07	15443.8	1.3	5233.0	1.1
Inverter-08	29067.1	1.3	4662.6	1.2
Inverter-09	6831.0	1.2	7535.9	1.0
Inverter-10	25311.0	1.1	9121.1	1.1
Inverter-11	33969.3	1.0	3819.8	1.0
Inverter-12	40980.6	1.1	4546.6	1.2
Inverter-13	17336.9	1.0	5722.0	1.5
Inverter-14	38946.1	1.2	4635.8	1.1
Inverter-15	10186.1	1.5	6903.4	1.4
Inverter-16	23420.2	1.2	7432.5	1.1
Inverter-17	42042.6	1.2	4347.0	1.1
Inverter-18	39023.8	1.1	5695.6	1.1
Inverter-19	45756.4	1.0	6527.6	1.1
Inverter-20	37640.6	1.1	6725.5	1.0
Inverter-21	31329.3	1.1	7391.6	1.2
Inverter-22	29849.1	1.1	5961.6	1.1
Inverter-23	45145.6	1.0	7714.2	1.1
Inverter-24	52253.1	1.1	6728.5	1.3

Table 9.10: **GFP and RFP expression of promoters J23100 and pBAD-insulated 24 inverter permutations on induction with 1mM L-arabinose.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for promoters J23100 and promoter pBAD insulated 24 inverter set in the absence of induction with L-arabinose for 14 hours. Data graphed in Figure 6.2.3.

ARABINOSE -				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	1939.4	1.2	42973.4	1.4
Inverter-02	4288.4	1.1	12092.4	1.3
Inverter-03	1817.8	1.2	25232.1	1.1
Inverter-04	3700.4	1.2	14206.0	1.2
Inverter-05	2306.6	1.2	18847.0	1.6
Inverter-06	2982.6	1.2	20486.3	1.7
Inverter-07	3010.0	1.1	50346.2	1.1
Inverter-08	4059.2	1.3	23369.6	1.9
Inverter-09	1209.4	1.1	36526.5	1.6
Inverter-10	2064.3	1.3	41578.0	1.5
Inverter-11	3634.3	1.2	16915.7	1.4
Inverter-12	4493.7	1.4	44726.6	1.4
Inverter-13	2380.1	1.4	22571.8	1.3
Inverter-14	3378.2	1.3	17595.2	1.1
Inverter-15	1551.5	1.1	64704.7	1.3
Inverter-16	2445.0	1.1	61465.6	1.2
Inverter-17	3490.6	1.3	29455.3	1.1
Inverter-18	2215.7	1.3	45977.9	1.5
Inverter-19	2662.4	1.1	40701.4	1.1
Inverter-20	2200.9	1.2	23833.9	1.3
Inverter-21	2099.2	1.2	68742.4	1.2
Inverter-22	4154.3	1.1	17676.8	1.1
Inverter-23	2606.4	1.2	53917.2	1.1
Inverter-24	4158.0	1.2	38753.8	1.3

Table 9.11: **GFP and RFP expression upon induction with 1mM L-arabinose of the 24 inverter set with all promoters (J23100, pBAD and pTet) insulated.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for the 24 inverter set with all promoters (J23100, pBAD and pet) insulated upon induction with L-arabinose for 14 hours. Data graphed in Figure 6.3.

ARABINOSE +				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	10185.07	1.88	2415.13	1.40
Inverter-02	34500.47	1.19	2070.15	1.31
Inverter-03	17653.51	1.34	221.05	1.48
Inverter-04	49117.25	1.04	1831.23	1.34
Inverter-05	37266.96	1.20	545.61	1.68
Inverter-06	59943.90	1.14	2338.39	1.26
Inverter-07	23243.47	1.06	1711.98	1.21
Inverter-08	39285.29	1.36	1199.51	1.21
Inverter-09	7893.40	1.46	3033.00	1.35
Inverter-10	28017.48	1.35	3781.72	1.25
Inverter-11	39056.02	1.48	1966.13	1.15
Inverter-12	36820.61	1.44	3440.93	1.49
Inverter-13	7906.04	1.08	2743.90	1.17
Inverter-14	29575.22	1.99	1901.21	1.12
Inverter-15	19794.25	1.70	2860.51	1.22
Inverter-16	38507.12	2.94	6992.46	1.53
Inverter-17	37678.19	1.09	2060.98	1.11
Inverter-18	36873.64	1.22	2552.04	1.21
Inverter-19	47606.47	1.11	2399.52	1.35
Inverter-20	35522.54	1.12	2095.72	1.16
Inverter-21	19547.17	1.80	3948.63	1.17
Inverter-22	40723.30	1.22	2623.13	1.36
Inverter-23	61546.59	1.24	4516.01	1.21
Inverter-24	43257.36	1.05	2917.08	1.26

Table 9.12: **GFP and RFP expression upon induction with 1mM L-arabinose of the 24 inverter set with all promoters (J23100, pBAD and pTet) insulated.** This table presents the geometric mean of the molecules of equivalent fluorescein (MEFL) for the 24 inverter set with all promoters (J23100, pBAD and pet) insulated in the absence of induction with L-arabinose for 14 hours. Data graphed in Figure 6.3.

ARABINOSE -				
Device ID	GFP		RFP	
	Mean	Std. Dev. of Mean (fold)	Mean	Std. Dev. of Mean (fold)
Inverter-01	2411.89	1.24	14403.73	1.35
Inverter-02	6216.97	1.15	10873.22	1.32
Inverter-03	3443.18	1.04	12738.14	1.50
Inverter-04	2904.28	1.10	11887.90	1.51
Inverter-05	3238.13	1.05	13044.01	1.49
Inverter-06	2519.01	1.07	23115.01	1.42
Inverter-07	2225.88	1.04	23684.65	1.28
Inverter-08	4347.49	1.30	11864.35	1.30
Inverter-09	1975.48	1.17	21472.77	1.37
Inverter-10	1953.31	1.15	20927.77	1.40
Inverter-11	4801.68	1.15	12945.03	1.18
Inverter-12	2980.43	1.11	23395.86	1.27
Inverter-13	2271.85	1.12	18817.94	1.65
Inverter-14	2977.22	1.09	20058.25	1.35
Inverter-15	1823.54	1.10	27262.14	1.17
Inverter-16	1933.86	1.04	42028.80	1.16
Inverter-17	2505.62	1.11	20829.12	1.31
Inverter-18	1987.06	1.11	24220.44	1.46
Inverter-19	2393.75	1.17	19343.59	1.22
Inverter-20	2943.30	1.10	16232.22	1.30
Inverter-21	1833.01	1.04	17184.15	1.53
Inverter-22	3163.14	1.22	14543.00	1.26
Inverter-23	2174.74	1.03	38395.38	1.29
Inverter-24	2697.06	1.21	27161.08	1.68

Chapter 10

Appendix B: PERL Script for DNA

Spacer Primer Design for MoClo-format promoters

```
#!/usr/bin/perl
#Swati Banerjee
#Conceived by Douglas Densmore
#Rev 1.0
#9/25/15
#Spacer Primer Generator"

#Minimal error checking
$numArgs = $#ARGV + 1;
if ($numArgs < 6 || $numArgs > 6) {
print "Required format is:\n";
print "-F <FusionSite> -S <Sequence File> -O <Output File>";
}
else
{
foreach $argnum (0 .. $#ARGV) {
```

```
if($ARGV[$argnum-1] eq "-F")
{$fusionSite = $ARGV[$argnum];}

if($ARGV[$argnum-1] eq "-S")
{$sequenceFile = $ARGV[$argnum];}

if($ARGV[$argnum-1] eq "-O")
{$outputFile = $ARGV[$argnum];}

}

print "Running for Fusion Site $fusionSite,
      Sequence File $sequenceFile, and Output File
      $outputFile \n";

open INFILE, $sequenceFile or die "Couldn't open file: $!";
$stringSequence = <INFILE>;
close INFILE;

#18 N's - this is the start of both the forward and reverse primer
$eighteenN = "NNNNNNNNNNNNNNNNNNNN";

#First 43bp of the sequence -
  together with the 18Ns will result in a 61bp
  forward primer
```

```
$firstFortyThree = substr($stringSequence, 0, 43);
#Create the forward primer
$forwardPrimer = $eighteenN . $firstFortyThree;

#Get the last 40bp; needed for the reverse primer
$lastForty = substr($stringSequence, -40);
# reverse the DNA sequence
$revcomp40 = reverse($lastForty);
# complement the reversed DNA sequence
    $revcomp40 =~ tr/ACGTacgt/TGCAtgca/;

#reverse the fusion site
$revFS = reverse($fusionSite);
$revFS =~ tr/ACGTacgt/TGCAtgca/;
#Create the reverse primer
$reversePrimer = $eighteenN . $revFS . $revcomp40;

$fpPrompt = "Forward Spacer Primer with FS
    $fusionSite upstream of promoter:";
$rpPrompt = "Reverse Primer with FS
    $fusionSite upstream of promoter:";

open (OUTFILE, ">", $outputFile);
print OUTFILE $fpPrompt;
print OUTFILE "\n";
print OUTFILE $forwardPrimer;
```

```
print OUTFILE "\n\n";
print OUTFILE $rpPrompt;
print OUTFILE "\n";
print OUTFILE $reversePrimer;
print OUTFILE "\n";
close OUTFILE;
}
```

Bibliography

- [1] Addgene. Addgene non-profit plasmid repository, 2016.
- [2] Christina M Agapakis and Pamela A Silver. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular BioSystems*, 5(7):704–713, 2009.
- [3] Hal Alper, Curt Fischer, Elke Nevoigt, and Gregory Stephanopoulos. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 102(36):12678–12683, 2005.
- [4] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol*, 2:2006 0028, 2006.
- [5] Narayana Annaluru, Héloïse Muller, Leslie A Mitchell, Sivaprakash Ramalingam, Giovanni Stracquadanio, Sarah M Richardson, Jessica S Dymond, Zheng Kuang, Lisa Z Scheifele, Eric M Cooper, et al. Total synthesis of a functional designer eukaryotic chromosome. *Science*, 344(6179):55–58, 2014.
- [6] Evan Appleton, Jenhan Tao, Traci Haddock, and Douglas Densmore. Interactive assembly algorithms for molecular cloning. *Nat. Methods*, 11:657–662, 2014.
- [7] Jacob Beal. Bridging the gap: a roadmap to breaking the biological design barrier. *Frontiers in bioengineering and biotechnology*, 2, 2015.
- [8] Jacob Beal, Ron Weiss, Douglas Densmore, Aaron Adler, Jonathan Babb, Swapnil Bhatia, Noah Davidsohn, Traci Haddock, Fusun Yaman, Richard Schantz, et al. Tasbe: A tool-chain to accelerate synthetic biological engineering. In *Proceedings of the 3rd International Workshop on Bio-Design Automation*, pages 19–21, 2011.
- [9] Jacob Beal, Ron Weiss, Fusun Yaman, Noah Davidsohn, and Aaron Adler. A method for fast, high-precision characterization of synthetic biology devices. Technical Report MIT-CSAIL-TR-2012-008, MIT, April 2012. Technical Report: MIT-CSAIL-TR-2012-008.

- [10] Lesia Bilitchenko, Adam Liu, Sherine Cheung, Emma Weeding, Bing Xia, Mariana Leguia, J Christopher Anderson, and Douglas Densmore. Eugene—a domain specific language for specifying and constraining synthetic biological parts, devices, and systems. *PloS one*, 6(4):e18882, 2011.
- [11] J. A. Brophy and C. A. Voigt. Principles of genetic circuit design. *Nat Methods*, 11(5):508–20, 2014.
- [12] Yizhi Cai. *GenoCAD: linguistic approaches to synthetic biology*. PhD thesis, Virginia Polytechnic Institute and State University, 2010.
- [13] S. Cardinale and A. P. Arkin. Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems. *Biotechnol J*, 7(7):856–66, 2012.
- [14] Gavin E Crooks, Gary Hon, John-Marc Chandonia, and Steven E Brenner. Weblogo: a sequence logo generator. *Genome research*, 14(6):1188–1190, 2004.
- [15] Thomas Crouzier. Researchers in the cloud. *Trends Biochem. Sci.*, 39:344–346, 2014.
- [16] J. H. Davis, A. J. Rubin, and R. T. Sauer. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res*, 39(3):1131–41, 2011.
- [17] Douglas Densmore, Joshua T Kittleson, Lesia Bilitchenko, Adam Liu, and J Christopher Anderson. Rule based constraints for the construction of genetic devices. In *ISCAS*, pages 557–560, 2010.
- [18] Densmore-CIDAR. AutomatedMoClo.esc evoware instructions for moclo assembly, 2016.
- [19] C. Engler, R. Kandzia, and S. Marillonnet. A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, 3(11):e3647, 2008.
- [20] C. Engler, M. Youles, R. Gruetzner, T. M. Ehnert, S. Werner, J. D. Jones, N. J. Patron, and S. Marillonnet. A golden gate modular cloning toolbox for plants. *ACS Synth Biol*, 3(11):839–43, 2014.
- [21] Shawn T Estrem, Tamas Gaal, Wilma Ross, and Richard L Gourse. Identification of an up element consensus sequence for bacterial promoters. *Proceedings of the National Academy of Sciences*, 95(17):9761–9766, 1998.
- [22] Shawn T Estrem, Wilma Ross, Tamas Gaal, ZW Susan Chen, Wei Niu, Richard H Ebright, and Richard L Gourse. Bacterial promoter architecture:

- subsite structure of up elements and interactions with the carboxy-terminal domain of the rna polymerase α subunit. *Genes & development*, 13(16):2134–2147, 1999.
- [23] Anthony C Forster and George M Church. Towards synthesis of a minimal cell. *Molecular systems biology*, 2(1):45, 2006.
- [24] Michal Galdzicki, Mandy Wilson, Cesar A Rodriguez, Matthew R Pocock, Ernst Oberortner, Laura Adam, Aaron Adler, J Christopher Anderson, Jacob Beal, Yizhi Cai, et al. Synthetic biology open language (sbol) version 1.1. 0. 2012.
- [25] Pamela K Geyer. The role of insulator elements in defining domains of gene expression. *Current opinion in genetics & development*, 7(2):242–248, 1997.
- [26] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, 3rd Hutchison, C. A., and H. O. Smith. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nat Methods*, 6(5):343–5, 2009.
- [27] Daniel G Gibson, Gwynedd A Benders, Cynthia Andrews-Pfannkoch, Evgeniya A Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B Stockwell, Anushka Brownley, David W Thomas, Mikkel A Algire, et al. Complete chemical synthesis, assembly, and cloning of a mycoplasma genitalium genome. *science*, 319(5867):1215–1220, 2008.
- [28] Daniel G Gibson, John I Glass, Carole Lartigue, Vladimir N Noskov, Ray-Yuan Chuang, Mikkel A Algire, Gwynedd A Benders, Michael G Montague, Li Ma, Monzia M Moodie, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *science*, 329(5987):52–56, 2010.
- [29] John I Glass, Nacyra Assad-Garcia, Nina Alperovich, Shibu Yooseph, Matthew R Lewis, Mahir Maruf, Clyde A Hutchison, Hamilton O Smith, and J Craig Venter. Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2):425–430, 2006.
- [30] Traci L Haddock, Douglas M Densmore, Evan Appleton, Swati Carr, Sonya Iverson, Monique De Freitas, Shawn Jin, Jake Awtry, Devina Desai, Thomas Lozanoski, et al. Bbf rfc 94: Type iis assembly for bacterial transcriptional units: A standardized assembly method for building bacterial transcriptional units using the type iis restriction enzymes bsai and bbsi. 2015.
- [31] Timothy S Ham, Zinovii Dmytriv, Hector Plahar, Joanna Chen, Nathan J Hillson, and Jay D Keasling. Design, implementation and practice of jbei-ice: an open source biological part registry platform and tools. *Nucleic acids research*, page gks531, 2012.

- [32] Karin Hammer, Ivan Mijakovic, and Peter Ruhdal Jensen. Synthetic promoter libraries—tuning of gene expression. *Trends in biotechnology*, 24(2):53–55, 2006.
- [33] Clyde A Hutchison, Scott N Peterson, Steven R Gill, Robin T Cline, Owen White, Claire M Fraser, Hamilton O Smith, and J Craig Venter. Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, 286(5447):2165–2169, 1999.
- [34] iGEM Foundation. Constitutive Promoter Family igem foundation, 1999.
- [35] iGEM Foundation. iGEM Registry of Standard Biological Parts igem foundation, 1999.
- [36] Sperotech Inc. Measuring molecules of equivalent FLuorochrome (MEF) using Sphero™ rainbow and ultra rainbow calibration particles rainbow beads, 2011.
- [37] Tecan Inc. Tecan freedom evo 150, 2016.
- [38] S. V. Iverson, T. L. Haddock, J. Beal, and D. M. Densmore. Cidar moclo: Improved moclo assembly standard and new e. coli part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synth Biol*, 5(1):99–103, 2016.
- [39] François Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, 3(3):318–356, 1961.
- [40] Peter Ruhdal Jensen and Karin Hammer. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Applied and environmental microbiology*, 64(1):82–87, 1998.
- [41] Woojin Jeong and Changwon Kang. Start site selection at lacuv5 promoter affected by the sequence context around the initiation sites. *Nucleic acids research*, 22(22):4667–4672, 1994.
- [42] Linda J Kahl and Drew Endy. A survey of enabling technologies in synthetic biology. *Journal of biological engineering*, 7(1):1, 2013.
- [43] Jason R Kelly, Adam J Rubin, Joseph H Davis, Caroline M Ajo-Franklin, John Cumbers, Michael J Czar, Kim de Mora, Aaron L Gliberman, Dileep D Monie, and Drew Endy. Measuring the activity of biobrick promoters using an in vivo reference standard. *Journal of biological engineering*, 3(1):4, 2009.
- [44] M. E. Lee, W. C. DeLoache, B. Cervantes, and J. E. Dueber. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol*, 4(9):975–86, 2015.

- [45] C. Lou, B. Stanton, Y. J. Chen, B. Munsky, and C. A. Voigt. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotechnol*, 30(11):1137–42, 2012.
- [46] Lance Martin, Austin Che, and Drew Endy. Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS One*, 4(11):e7569, 2009.
- [47] Wenmao Meng, Tamara Belyaeva, Nigel J Savery, Stephen JW Busby, Wilma E Ross, Tamas Gaal, Richard L Gourse, and Mark S Thomas. Up element-dependent transcription at the escherichia coli rrnB p1 promoter: positional requirements and role of the rna polymerase α subunit linker. *Nucleic acids research*, 29(20):4166–4178, 2001.
- [48] Graeme Mitchison. The regional rule for bacterial base composition. *Trends in Genetics*, 21(8):440–443, 2005.
- [49] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C Stanton, and Christopher A Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–253, 2012.
- [50] V. K. Mutalik, J. C. Guimaraes, G. Cambray, C. Lam, M. J. Christoffersen, Q. A. Mai, A. B. Tran, M. Paull, J. D. Keasling, A. P. Arkin, and D. Endy. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods*, 10(4):354–60, 2013.
- [51] Nikolai Naryshkin, Andrey Revyakin, Younggyu Kim, Vladimir Mekler, and Richard H Ebright. Structural organization of the rna polymerase-promoter open complex. *Cell*, 101(6):601–611, 2000.
- [52] Patricia Niland, Rolf Hühne, and Benno Müller-Hill. How arac interacts specifically with its target dnas. *Journal of molecular biology*, 264(4):667–674, 1996.
- [53] Ernst Oberortner, Swapnil Bhatia, Erik Lindgren, and Douglas Densmore. A rule-based design specification language for synthetic biology. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 11(3):25, 2014.
- [54] Ernst Oberortner and Douglas Densmore. Web-based software tool for constraint-based design specification of synthetic biological systems. *ACS synthetic biology*, 4(6):757–760, 2014.
- [55] AS Orekhova and PM Rubtsov. Bidirectional promoters in the transcription of mammalian genomes. *Biochemistry (Moscow)*, 78(4):335–341, 2013.
- [56] Stephanie Paige, Sonya Iverson, Aaron Heuckroth, Swati Carr, Traci Haddock, and Douglas Densmore. The boston university center of synthetic biology (cosbi) ice repository.

- [57] György Pósfai, Guy Plunkett, Tamás Fehér, David Frisch, Günther M Keil, Kinga Umenhoffer, Vitaliy Kolisnychenko, Buffy Stahl, Shamik S Sharma, Monika De Arruda, et al. Emergent properties of reduced-genome escherichia coli. *Science*, 312(5776):1044–1046, 2006.
- [58] P. E. Purnick and R. Weiss. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*, 10(6):410–22, 2009.
- [59] Lei Qi, Rachel E Haurwitz, Wenjun Shao, Jennifer A Doudna, and Adam P Arkin. Rna processing enables predictable programming of gene expression. *Nature biotechnology*, 30(10):1002–1006, 2012.
- [60] Nicholas Roehner. *Technology mapping of genetic circuit designs*. PhD thesis, The University of Utah, 2014.
- [61] Nicholas Roehner and Douglas Densmore. Double dutch: A tool for designing libraries of variant metabolic pathways.
- [62] Wilma Ross, Alexander Ernst, and Richard L Gourse. Fine structure of e. coli rna polymerase-promoter interactions: α subunit binding to the up element minor groove. *Genes & development*, 15(5):491–506, 2001.
- [63] A. Sarrion-Perdigones, E. E. Falconi, S. I. Zandalinas, P. Juarez, A. Fernandez-del Carmen, A. Granell, and D. Orzaez. Goldenbraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One*, 6(7):e21622, 2011.
- [64] Christina D Smolke. Building outside of the box: igem and the biobricks foundation. *Nature biotechnology*, 27(12):1099–1102, 2009.
- [65] Martin Christen Frølund Thomsen and Morten Nielsen. Seq2logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic acids research*, 40(W1):W281–W287, 2012.
- [66] Prashant Vaidyanathan, Bryan S Der, Swapnil Bhatia, Nicholas Roehner, Ryan Silva, Christopher A Voigt, and Douglas Densmore. A framework for genetic logic synthesis. *Proceedings of the IEEE*, 103(11):2196–2207, 2015.
- [67] Kimberly A Walker and Robert Osuna. Factors affecting start site selection at the escherichia coli fis promoter. *Journal of bacteriology*, 184(17):4783–4791, 2002.

- [68] E. Weber, C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet. A modular cloning system for standardized assembly of multigene constructs. *PLoS One*, 6(2):e16765, 2011.
- [69] Wu Wei, Vicent Pelechano, Aino I Järvelin, and Lars M Steinmetz. Functional consequences of bidirectional promoters. *Trends in Genetics*, 27(7):267–276, 2011.
- [70] Bing Xia, Swapnil Bhatia, Ben Bubenheim, Maisam Dadgar, Douglas Densmore, and J Christopher Anderson. Developers and users guide to clotho v2.0 a software platform for the creation of synthetic biological systems. *Methods Enzymol*, 498:97–135, 2011.
- [71] Enoch Yeung, Andrew Ng, Jongmin Kim, Zachary Z Sun, and Richard M Murray. Modeling the effects of compositional context on promoter activity in an e. coli extract based transcription-translation system. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 5405–5412. IEEE, 2014.

Curriculum Vitae

Name **Swati B. Carr**

Molecular Biology, Cell Biology & Biochemistry Program. 36 Cummington Mall, Boston, MA 02215

Education **Boston University**, PhD in Synthetic Biology (Expected: September 2016) Thesis advisor: Douglas M. Densmore

University of Rhode Island, MS in Cell & Molecular Biology (Concentration: Microbial Genetics), September 2006 – May 2009.

Bangalore University, B.Sc, Biotechnology, Chemistry, Zoology September 2002 – May 2005.

Research Skills

Molecular Biology

- DNA/RNA isolation, PCR, qPCR, Site-directed mutagenesis
- Standardized DNA assembly methods: Modular Cloning (MoClo), Golden Gate, Gibson Assembly, Gateway, BioBricks
- standard molecular biology techniques including gel electrophoresis and sub-cloning
- Protein Assays: Chromatin Immunoprecipitation (ChIP), SDS-PAGE, Western Blotting, ELISA

Flow Cytometry

Mammalian Biology

- in vitro culture of primary and immortalized cell lines
- in vivo assays in mouse models

Bioinformatics Tools

- Alignment Tools: ClustalW, MEGA, NCBI BLAST

Awards

2015

- SynBERC Travel Award to attend Gordon Research Conference in Synthetic Biology (2015)
- Travel Scholarship to International Workshop for Bio-Design Automation (IWBDAA) (2015)

2014

- National Science Foundation Global Change Initiative: Education and Research (GLACIER) GK-12 Fellowship (2014-2015)

2013

- Best poster award at the UK-US Partnerships in Synthetic Biology (2013)
- Travel Scholarship to International Workshop for Bio-Design Automation (IWBD A) (2013)

2012

- International Genetically Engineered Machine Competition (iGEM) Gold Medal (2012)

2009

- Boston University Deans Fellowship (2009)

- Publications*
1. Swati B. Carr, Jacob Beal, Douglas Densmore. DNA component order independence and precise expression control in bacteria using a novel insulated promoter screening technique. (manuscript in preparation)
 2. Swati B. Carr, Luis Ortiz, Rachel Smith, Douglas Densmore. Automated Liquid Handling Assembly of Modular DNA Devices. *Journal of Visual Experiments*. (manuscript in review)
 3. Traci L. Haddock, Douglas M. Densmore, Evan Appleton, Swati Carr, Sonya Iverson, Monique De Freitas, Shawn Jin et al. "BBF RFC 94: Type IIS Assembly for Bacterial Transcriptional Units: A Standardized Assembly Method for Building Bacterial Transcriptional Units Using the Type IIS Restriction Enzymes BsaI and BbsI." (2015).
 4. B. Yordanov, E. Appleton, R. Ganguly, E. A. Gol, S. B. Carr, S. Bhatia, T. Haddock, C. Belta, and D. Densmore, Experimentally driven verification of synthetic biological circuits, in *Proceedings of the Conference on Design, Automation and Test in Europe*, Dresden, Germany, 2012, pp. 236-241.
 5. Leatham, Mary P., et al. "Precolonized human commensal *Escherichia coli* strains serve as a barrier to *E. coli* O157: H7 growth in the streptomycin-treated mouse intestine." *Infection and immunity* 77.7 (2009): 2876-2886.
 6. Terragni, Jolyon, et al. "The E-box binding factors Max/Mnt, MITF, and USF1 act coordinately with FoxO to regulate expression of proapoptotic and cell cycle control genes by phosphatidylinositol 3-kinase/Akt/glycogen

synthase kinase 3 signaling.” *Journal of Biological Chemistry* 286.42 (2011): 36215-36227.