ABSTRACT

AFROZ, TALIMAN. Understanding and Engineering Individuality in Bacterial Sugar Utilization. (Under the direction of Dr. Chase L. Beisel).

Bacteria, like humans, can exhibit ‘individuality’ when faced with similar choices. The presence of positive and negative feedback in regulatory systems plays a crucial role in dictating ‘individuality’ in bacteria. Interestingly, positive and negative feedback loops are also present in sugar utilization pathways ubiquitous to virtually all microorganisms. The bacterial sugar utilization pathways encode regulators, transporters and catabolic enzymes that are induced by their cognate sugars. The regulators control the expression of the pathway components; transporters import more sugar conferring positive feedback and the enzymes break down the sugar imparting negative feedback. Because of scientific and biotechnological implications, sugar utilization pathways in the model bacterium *Escherichia coli* have been studied for many decades. However, most of the work done so far involved bulk characterization techniques that fail to capture unique behaviors observable only at the single-cell level. The very few single-cell studies available were focused on only two pathways. More importantly, the studies neglected sugar catabolism, which is an integral part of the natural sugar utilization pathways. Therefore, it remains unclear how the myriad of natural sugar utilization pathways respond to sugars at the single-cell level. This thesis work addresses this gap by investigating the single-cell response of *E. coli* to eight different sugars using a combination of experimental and computational approaches. The response of *E. coli* to sugars was remarkably diverse, with negative feedback due to catabolism playing a crucial role. Building on these insights, we engineered cells to convert sharp bimodal responses of natural sugar utilization pathways into linear, titratable responses. The motivation behind the work was to use the natural inducible sugar utilization pathways as ‘in-built’ titratable systems in non-model organisms. Using the L-arabinose and D-xylose utilization pathways as model systems, we showed that each modification came with a trade-off. Finally, as an extension to this work we showed how the inadvertent presence of inducer in the medium could affect the response properties such as linearity and dynamic range. Overall, this work provides a few significant insights (a) feedback loops in bacterial sugar utilization are one of the crucial factors that help the cells attain ‘individuality’ or cope under fluctuating nutrient
conditions (b) contrary to the prevailing assumptions, response of these ubiquitous utilization pathways can be extremely complex, with potential effects on how the cells can be used for various biotechnological applications (c) the endogenous sugar utilization pathways, with modifications, can be co-opted as titratable systems especially in non-model microorganisms. As part of future work, the strength of the feedbacks loops could be tuned to understand their effects on the transition rates between induced and uninduced states. Also, the single-cell analyses could be extended to microorganisms beyond *E. coli* to verify the generality of the observed diverse responses.
Understanding and Engineering Individuality in Bacterial Sugar Utilization

by
Taliman Afroz

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Chemical Engineering

Raleigh, North Carolina
2015

APPROVED BY:

_______________________________  ______________________________
Dr. Chase L. Beisel             Dr. Robert Kelly
Committee Chair

_______________________________  ______________________________
Dr. Amy Grunden                 Dr. Jason Haugh
DEDICATION

This thesis is dedicated to my parents
BIOGRAPHY

Taliman Afroz was born on February 6, 1984 in Dhaka, Bangladesh. After graduating from Maple Leaf International School in 2001, she attended Bangladesh University of Engineering & Technology (BUET) for her undergraduate studies. In 2007, she graduated with her Bachelor of Science in Chemical Engineering from BUET. She worked as a junior lecturer in the department for two years. In the Fall semester of 2009, she joined the Department of Chemical & Biomolecular Engineering at North Carolina State University as a graduate student. After completing her masters degree in two years, she joined the Beisel group in the Spring semester of 2012 to pursue her PhD.
ACKNOWLEDGMENTS

First I would like to thank my advisor Dr Chase L. Beisel for his constant guidance and inspiration. I will always be thankful to him for believing in me and in my abilities. Also, I feel blessed to have such smart, helpful group members who always provided useful insights or constructive criticisms about my work.

I gratefully acknowledge the help I got from all the staff of the Chemical & Biomolecular Engineering Department. I would specially like to thank Sandra Bailey for being so nice and helpful.

Finally, I would like to thank my parents, brothers and my husband, Rashed, for their unconditional love and support. Last but not the least I would like to thank all those ordinary people of Bangladesh whose tax money helped me to get an almost free education. I will forever be indebted to them.
# TABLE OF CONTENTS

| LIST OF TABLES  | - | - | - | - | - | - | - | ix |
| LIST OF FIGURES | - | - | - | - | - | - | - | xi |

**Chapter 1  Introduction and Background**  
1.1 Individuality in bacteria  
1.2 Bacterial sugar utilization  
1.3 Individuality in bacterial sugar utilization?  
1.4 Overview of the thesis  
References  

**Chapter 2 Understanding and Exploiting Feedback in Synthetic Biology**  
2.1. Introduction  
2.2. Modes of feedback  
2.2a Transcriptional regulation  
2.2b Post-transcriptional regulation  
2.2c Post-translational regulation  
2.2d Cell-cell interactions  
2.3. Properties of negative feedback  
2.3a Dynamics: Shorter response time  
2.3b Steady-state: Linearized response  
2.3c Cell-cell variability: Noise reduction  
2.4. Properties of positive feedback  
2.4a Dynamics: Longer response time  
2.4b Steady state: Ultrasensitivity  
2.4c Cell-cell variability: Increased noise and bistability  
2.5. Properties of combined negative and positive feedback  
2.5a Multi-negative feedback loops  

References  

Chapter 1  Introduction and Background  
1.1 Individuality in bacteria  
1.2 Bacterial sugar utilization  
1.3 Individuality in bacterial sugar utilization?  
1.4 Overview of the thesis  
References  

Chapter 2 Understanding and Exploiting Feedback in Synthetic Biology  
2.1. Introduction  
2.2. Modes of feedback  
2.2a Transcriptional regulation  
2.2b Post-transcriptional regulation  
2.2c Post-translational regulation  
2.2d Cell-cell interactions  
2.3. Properties of negative feedback  
2.3a Dynamics: Shorter response time  
2.3b Steady-state: Linearized response  
2.3c Cell-cell variability: Noise reduction  
2.4. Properties of positive feedback  
2.4a Dynamics: Longer response time  
2.4b Steady state: Ultrasensitivity  
2.4c Cell-cell variability: Increased noise and bistability  
2.5. Properties of combined negative and positive feedback  
2.5a Multi-negative feedback loops  

References
### Chapter 6 Future work

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Single-cell responses of <em>E. coli</em> to mixtures of sugars</td>
<td>178</td>
</tr>
<tr>
<td>6.2 Extension of single-cell studies to sugar utilization pathways of microorganisms beyond <em>E. coli</em></td>
<td>179</td>
</tr>
<tr>
<td>6.3 Apply the rules derived for using natural sugar utilization pathways as titratable systems for non-model microorganisms</td>
<td>180</td>
</tr>
<tr>
<td>6.4 Tuning catabolism using CRISPER-cas system</td>
<td>180</td>
</tr>
<tr>
<td>References</td>
<td>182</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<p>| Table S3.1 | Parameters of the model | - | - | - | - | - | 90 |
| Table S3.2 | Dimensionless variables | - | - | - | - | - | 91 |
| Table S3.3 | Overview of utilization pathways with inducibly expressed transporters and catabolic enzymes in <em>E. coli</em> | - | - | - | - | - | 100 |
| Table S3.4 | Doubling times in the presence or absence of sugar. Doubling times were measured for the indicated strain growing in exponential phase (ABS$_{600}$ of 0.01 – 0.3) in M9 glycerol without (-S) or with (+S) 1 mM of the indicated sugar | - | - | - | - | - | 101 |
| Table S3.5 | Parameters for the model of sugar utilization with a sensory activator or repressor. The potential biological ranges for each parameter were calculated based on order-of-magnitude estimates for each ratio | - | - | - | - | - | 102 |
| Table S3.6 | Strains used in this work | - | - | - | - | - | 104 |
| Table S3.7 | Plasmids used in this work | - | - | - | - | - | 105 |
| Table S3.8 | Oligonucleotides used in this work | - | - | - | - | - | 106 |
| Table 4.1 | Response metrics for each alteration of the L-arabinose utilization pathway in <em>E. coli</em> | - | - | - | - | - | 126 |
| Table S4.1 | Parameters of the model | - | - | - | - | - | 138 |
| Table S4.2 | Dimensionless parameters | - | - | - | - | - | 139 |
| Table S4.3 | Values of the dimensionless parameters | - | - | - | - | - | 141 |
| Table S4.4 | Measured growth rates for each strain | - | - | - | - | - | 148 |
| Table S4.5 | Strains used in this study | - | - | - | - | - | 149 |
| Table S4.6 | Plasmids used in this study | - | - | - | - | - | 150 |</p>
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4.7</td>
<td>Oligonucleotides used in this study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>152</td>
</tr>
<tr>
<td>S5.1</td>
<td>Strains used in this study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>172</td>
</tr>
<tr>
<td>S5.2</td>
<td>Plasmids used in this study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>172</td>
</tr>
<tr>
<td>S5.3</td>
<td>Oligonucleotides used in this study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>173</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 1.1</td>
<td>Carbon catabolite repression in <em>Escherichia coli</em></td>
<td>09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Repression of <em>xyl</em> operon in presence of L-arabinose</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Modes of feedback in microorganisms</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Contrasting characteristics of negative and positive feedback</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Combining negative and positive feedback</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Current and future applications of feedback in synthetic biology</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Inducible sugar utilization</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Varying single-cell responses to different sugars in <em>E. coli</em></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>A simple model predicts varying responses from coupling inducible transport and inducible catabolism</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Single-cell responses in the absence of sugar catabolism</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Impact of endogenous biosynthesis of the sugar on the predicted response</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Extent of hysteresis for the D-xylose and L-arabinose utilization pathways with intact or disabled sugar catabolism</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure S3.1</td>
<td>Measuring pathway expression in single cells by flow cytometry</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure S3.2</td>
<td>Additional response curves for WT MG1655 cells</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure S3.3</td>
<td>Behaviors predicted by the deterministic model for unequal dilution rates of the sugar ($D_1 = 0.74$), transporter ($D_2 = 0.74$), and enzyme ($D_3 = 0.27$)</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure S 3.4</td>
<td>Similar behaviors predicted when the regulator functions as a sensory repressor</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure S3.5  Response curves for ΔxylAB and ΔaraBAD strains following excision of the cat resistance cassette - - - - - - - 97

Figure S3.6  Impact of model parameters on the width of the bifurcation region in the absence of catabolism - - - - - - - 98

Figure S3.7  The L-arabinose utilization pathway exhibits limited memory of previous conditions, but only in the presence of catabolism - - - - 99

Figure 4.1  Co-opting sugar utilization pathways for titratable control of gene expression - - - - - - - 111

Figure 4.2  A simple mathematical model predicts trade-offs when altering the pathway structure - - - - - - - 115

Figure 4.3  Probing alterations to the L-arabinose utilization pathway in E. coli 119

Figure 4.4  Effect of cell density in the presence or absence of sugar catabolism 122

Figure 4.5  Linearizing the response to D-xylose - - - - - - 124

Figure S4.1  The L-arabinose and D-xylose utilization pathways in E. coli - 143

Figure S4.2  Overview of flow cytometry analysis - - - - - - 144

Figure S4.3  Impact of incubation time on the response to L-arabinose - - 145

Figure S4.4  Predicted response when only one transporter is present - - 146

Figure S4.5  Response dynamics in the presence or absence of L-arabinose catabolism - - - - - - - 147

Figure 5.1  Mathematical modeling of residual inducer on titratable systems - 161

Figure 5.2  Impact of residual inducer on titratable systems in E. coli - - 164

Figure 5.3  Impact of residual inducer on an “all-or-none” system in E. coli - 166

Figure S5.1  Performance metrics based on residual inducer - - - 174
Figure S5.2. Predicted impact of residual inducer for varying values of the dynamic range or the half-maximal inducer concentration - - - - 175

Figure S5.3 Impact of pre-incubating residual L-arabinose on the response parameters - - - - - - - - 176
Chapter 1
Introduction and Background
1.1 Individuality in bacteria

Bacteria undergo spontaneous mutation to cope with heterogeneous environments but cells can also exhibit phenotypic variability under homogeneous environmental conditions (Veening et al., 2008). Non-genetic individuality refers to the tendency of bacteria to exhibit cell-to-cell variability in an isogenic population under uniform conditions. It is commonly believed to be a complex function of noise and the structure of the regulatory and biochemical networks (Becskei and Serrano, 2000; Davidson and Surette, 2008). From an evolutionary point of view, this kind of population heterogeneity is often thought to be used by the cells as a technique to maximize fitness under extreme or unpredictable conditions (King and Masel, 2007).

One of the classic examples of individuality is the existence of ‘persister’ cells. Bigger et. al showed that when *Streptococcus pyogenes* was treated with penicillin, a majority of the cells died while a small fraction survived (Bigger, 1944). When this small surviving population was grown and treated with penicillin, again most of the cells died leaving behind the resistant subpopulation called ‘persisters.’ This suggested that the subpopulation did not arise because they acquired resistance genes (Van Hoek et al., 2011) or mutations (Woodford and Ellington, 2007). In addition to *S. pyogenes*, ‘persisters’ have been observed in various bacterial species (Lewis, 2008; Woodford and Ellington, 2007). It was hypothesized for years that as antibiotics usually target actively growing cells, persisters escaped death by either growing slowly or not growing at all. The hypothesis was not confirmed until Balaban et. al used single-cell analysis to study persister cells in *Escherichia coli* (Balaban et al., 2004). They grew the GFP expressing cells in microfluidic device and using time-lapse microscopy followed the growth of each bacterium forming colonies. Intriguingly, they showed that the cells growing slowly were present even before they were treated with antibiotic. One possible explanation for this is that population heterogeneity is used by the cells as a bet-hedging technique to survive under unpredictable conditions.

Another well studied example of individuality in bacteria is the tendency of the gram-positive bacterium *Bacillus subtilis* to become ‘competent’ under nutrition deficient conditions (Dubnau and Losick, 2006). Competence refers to the ability of cells to uptake
DNA from the environment and integrate it into their genome (Chen et al., 2005). However, only a small fraction of cells from a genetically homogenous culture become competent under stress. Studies showed that competence is a transient and probabilistic phenomenon governed by the transcriptional regulator ComK (Hahn et al., 1996). The regulation of ComK is complex with a number of repressors and activators affecting its expression. Using time-lapse fluorescence microscopy for single-cell analysis, Suel et al demonstrated that the transient nature of competence is due to the fast positive and slow negative feedback affecting the concentration of ComK (Süel et al., 2006). ComK is autoregulated via a positive feedback loop (Smits et al., 2005) and noise triggers the change from vegetative to competent state. However, ComK also suppresses ComS (Hahn et al., 1994). As ComS represses ComK degradation by MecA complex (Ogura et al., 1999), an indirect negative feedback is thus acting on ComK. This relatively slow negative feedback helps the competent cells to escape competence and return to vegetative state. *Bacillus subtilis* also has tendency to form spores under nutrient deficient conditions. The sporulation is regulated by the transcription factor Spo0A. Using flow cytometry, Chung et al showed that a genetically homogenous group of cells form population heterogeneity where some cells turn on the master regulator for spore formation (Spo0A-ON cells) while others turn it off (Spo0A-OFF cells) (Chung et al., 1994). Similar to competence, sporulation is also thought to be affected by the presence of positive and negative feedback regulatory loops (Chung et al., 1994; Veening et al., 2005). From these examples a few things can be observed: (i) individuality is a common and intriguing phenomenon in bacterial systems, (ii) single-cell analysis is an extremely useful tool for analyzing individuality, and (iii) positive and negative feedback in the regulatory systems play a crucial role in dictating individuality in bacteria.

### 1.2 Bacterial sugar utilization

Interestingly, positive and negative feedback loops are also present in sugar utilization pathway ubiquitous to virtually all microorganisms (Carlson, 1987; Gunsalus et al., 1955). Nutrient utilization is essential for the growth and survival of all forms of life in this world. For microorganisms, usually there are dedicated pathways and mechanisms to recognize,
uptake and utilize nutrients from the environment. Bacteria, like other microorganisms, encode numerous sugar utilization pathways to uptake and utilize sugars as carbon and energy sources. In addition to survival, nutrient utilization is important for processes such as virulence (Rojo, 2010), biofilm formation (Chai et al., 2012; Sadykov et al., 2011), DNA uptake (Lo Scrudato and Blokesch, 2012) and cell-cell communication (Pereira et al., 2012). The colonization of microorganisms in human gut is also influenced by the presence and activity of carbon sources (Blaut and Clavel, 2007; Chang et al., 2004). Some recent studies have linked the activities and composition of the gut microbiome to obesity (DiBaise et al., 2012), diabetes (Gomes et al., 2014; Musso et al., 2010) and neurological diseases (Cryan and Dinan, 2012; Hemarajata and Versalovic, 2013).

Moreover, inducible sugar utilization pathways have applications. They are used as standard tools for inducible expression systems to express recombinant proteins (Sodoyer, 2004), regulate components in metabolic pathways (Nevoigt, 2008; Zhang et al., 2012) and investigate characteristics of novel genes and proteins in biochemical studies (Guzman et al., 1995). Temme et al., for example, refactored the nitrogen fixing gene cluster from Klebsiella oxytoca to simplify complex genetics. In doing so, they used inducible promoters to optimize the activities of the refactored parts (Temme et al., 2012). Also, knowledge about sugar utilization pathways is of paramount significance when cells are used for metabolic engineering purposes (Aristidou and Penttilä, 2000). In the last few decades, there has been increasing interest in utilizing microorganisms to produce biofuels, chemicals and other valuable end products (Zaldívar et al., 2001). Plant biomass, which contains mixtures of sugars, is a common starting material for most all these processes (Saha, 2003).

Inducible sugar utilization pathways consist of regulators, transporters and catabolic enzymes. The sugar induces the expression of these pathway components. The regulators control the expression of the pathway components; the transporters import sugars into the cells forming a positive feedback while the enzymes break down the sugar, imparting negative feedback. Even though bacterial sugar utilization has been widely studied, little is known how these feedback loops affect the response of bacteria, at the single-cell level, to different sugars. Could these feedback loops also give rise to ‘individuality’ in bacterial sugar
utilization? Answers to this question are crucial because individuality could potentially redefine, for example, how we engineer bacteria to convert lignocellulosic biomass into useful end products. Non-uniform response of cells to sugars would pose a bottleneck in metabolic engineering by inactivating a portion of the microbial population, thereby reducing product conversion rates. Obtaining knowledge about how each cell respond to the various sugars is, therefore, an essential step in fully understanding how microorganisms interact with the environment and is critical in engineering strains or using these pathways for diverse industrial applications.

1.3 Individuality in bacterial sugar utilization?

As mentioned above, bacterial sugar utilization has been studied for many decades. The model pathway, lac operon in Escherichia coli is, in fact, used in microbiology textbooks as a common example of bacterial gene regulation (Alberts et al., 2002). A majority of the studies, done using bulk characterization techniques, assuming uniform responses across a genetically homogenous population for varying sugar concentrations. For example, Kaplan et. al studied the effects of cAMP and sugar concentrations on the expression level of the pathway components, assuming uniform responses at all sugar concentrations (Kaplan et al., 2008). Again, to investigate the factors affecting the expression of lac operon, Kuhlman et. al used population average measurements (Kuhlman et al., 2007). Limited modeling work done on the lac operon, without experimental validation, also predicts uniform response (Krishna et al., 2007; Savageau, 2011). However, the coupled positive and negative feedback loops in bacterial sugar utilization could potentially give rise to individuality that the bulk studies would fail to capture. This implies the importance of using single-cell analyses for encapsulating the unique and important single-cell behaviors.

The very few single-cell studies revealed that, for some sub-saturating inducer concentrations, the cells bifurcate into fully induced and fully uninduced subpopulations. One critical feature of the single-cell studies was a lack of sugar consumption either through the use of a non-hydrolyzable inducer (Cohn and Horibata, 1959; Novick and Weiner, 1957) or through the deletion of the catabolic genes (Siegele and Hu, 1997). In the classic single-
cell study Novick *et. al* used the non-metabolizable lactose analogue thiomethyl-β-D-galactoside (TMG) to induce *E. coli* and got the bimodal all-or-none response (Novick and Weiner, 1957). Decades later, fluorescence microscopy was used to confirm the all-or-none response of *E. coli* to intermediate TMG concentrations (Ozbudak *et al.*, 2004). When one cell is stochastically induced, it produces the LacY permease which, in turn, brings in more inducer forming a positive feedback. The positive feedback drives the cell to full induction (Rao and Koirala, 2014). A similar all-or-none response was also observed for L-arabinose utilization pathway using cells that cannot catabolize arabinose (Khlebnikov *et al.*, 2001; Megerle *et al.*, 2008; Morgan-Kiss *et al.*, 2002).

Most importantly, the single-cell studies were limited to only two utilization pathways-*lac* operon and L-arabinose utilization pathway in *E. coli*. However, this model bacterium can grow on a large number of sugars and the method of induction and regulation of the pathway components vary with the inducing sugar. L-arabinose utilization pathway, for example, is induced by L-arabinose whereas the *lac* operon in *E. coli* is induced by the metabolic intermediate allolactose instead of the sugar lactose (Wheatley *et al.*, 2013). Again, the regulation of L-arabinose utilization pathway is complex compared to the *lac* operon (Schleif, 2010). AraC, the dual regulator of the L-arabinose utilization pathway, uses DNA looping to repress the *ara* operon in absence of L-arabinose but activates the operon in presence of the sugar (Lobell and Schleif, 1990). Also, the expression of AraC is negatively autoregulated (Lee *et al.*, 1981). Regulation of the *lac* operon, in comparison, is simpler with LacI repressing the genes to transport and catabolize lactose in absence of the sugar. Furthermore, even though the L-arabinose and D-xylose utilization pathways have similar structures (Song and Park, 1997), D-xylose utilization pathway is positively regulated by XylR and, unlike AraC, XylR uses DNA looping for activation of the pathway transporters and enzymes (Ni *et al.*, 2013). Regulation of the genes for L-rhamnose transport and catabolism, on the other hand, requires a regulatory cascade *rhaSR* (Tobin and Schleif, 1990). When L-rhamnose is present, basal level of RhaR binds to it and activates the *rhaSR* operon. As *rhaSR* is expressed, RhaS activates expression of *rhaT* and *rhaBAD* to transport and catabolize L-rhamnose respectively (Egan and Schleif, 1993).
Additionally, unlike the pathways discussed so far, there are pathways where the transport and catabolism involve two different systems. The gluconate utilization pathway is one such example. Gluconate is one of the preferred carbon sources of *Escherichia coli* and there are two routes, GntI and GntII for the transport and initial catabolism of D-gluconate (Bausch et al., 2004; Izu et al., 1997). Similar to LacI, GntR is a repressor and represses expression of the GntI pathway in absence of gluconate. Again, the activator of GntII pathway can act as the repressor for GntI pathway (Tsunedomi et al., 2003). While bacteria acquire most of the sugars from the surrounding, there are sugars which are endogenously produced inside the cells. N-acetylglucosamine and D-galactose are two such examples (Vogler et al., 1989; Weickert and Adhya, 1993a). N-acetylglucosamine, GlcNAc is an important building block for the peptidoglycan of bacterial cells and also important source of carbon and nitrogen for bacteria. The *glmUS* operon encodes the biosynthetic genes to convert the glycolytic product fructose-6-P into UDP-GlcNAc for synthesizing peptidoglycan and lipopolysaccharide (Mengin-Lecreulx and van Heijenoort, 1993; Walker et al., 1984). In presence of GlcNAc, the cells attain a balance between biosynthesis and catabolism using the transcriptional factor, NagC. NagC acts as an activator for the catabolic genes and as a repressor for the biosynthetic genes (Plumbridge, 1995; Plumbridge and Kolb, 1993). Similar to N-acetylglucosamine, D-galactose is also used as a building block for the cells and is an inadvertent by-product of the Leloir pathway (Holden et al., 2003; Weickert and Adhya, 1993a). The regulation of D-galactose utilization pathway is extremely complex with multiple regulators, transporters and enzymes (Kalckar, 1971; Semsey et al., 2007; Weickert and Adhya, 1993).

Cross regulation between pathways is also common in bacteria. One classic example is how the presence of D-glucose can repress expression of the genes required to metabolize lactose when bacteria are grown on mixtures of D-glucose and lactose (Loomis and Magasanik, 1967). The phenomenon is referred to as carbon catabolite repression (CCR) where the presence of a preferred carbon source (i.e. D-glucose for *E. coli*) restricts the cells from simultaneous utilization of other sugars present (Liu et al., 2005). The mechanism of catabolite repression is complex. The major components of CCR are: the transcriptional
activator CRP (cyclic AMP (cAMP) receptor protein), also known as catabolite gene activator protein (CAP); cAMP, the signaling molecule for CAP; adenylate cyclase and the IIA component of the D-glucose specific phosphoenolpyruvate phosphotransferase systems (PTS), EIIA\textsuperscript{Glc} (Görke and Stülke, 2008). Figure 1.1 summarizes the carbon catabolite mechanism in \textit{E. coli}. When EIIA\textsuperscript{Glc} is dephosphorylated, it binds and inactivates the metabolic enzymes and transporters of the secondary carbon source, such as lactose in a mixture of D-glucose and lactose. This mechanism is called inducer exclusion (Hogema et al., 1999). In its phosphorylated state, EIIA\textsuperscript{Glc} binds and activates adenylate cyclase (AC) forming cAMP. cAMP then binds to CRP and the CRP-cAMP complex then activates the expression of catabolic operons. The phosphorylation of EIIA\textsuperscript{Glc} is mainly dictated by the ratio of phosphoenolpyruvate to pyruvate in the cells (Hogema et al., 1998). The contribution of CRP-cAMP complex to CCR is not straightforward. Earlier it was believed that presence of D-glucose and cAMP is inversely correlated, higher D-glucose concentrations leading to lower levels of cAMP. However, it was found that the levels of cAMP is similar in cells growing in D-glucose or in lactose (Inada et al., 1996). In addition to activating the catabolic operon required for lactose metabolism, CRP-cAMP seems to activate transcription of the major D-glucose transporter gene \textit{ptsG} (Kimata et al., 1997). Interestingly, You \textit{et. al} recently showed that cAMP signaling is not centered on carbon metabolism but is rather dependent on the balance between proteomic resources and metabolic demands (You et al., 2013).

Similar to carbon catabolite repression by D-glucose, the hierarchical utilization of sugar consumption also exists among other sugars (Desai and Rao, 2010; Groff et al., 2012). When \textit{E. coli} is grown in a mixture of L-arabinose and D-xylose, \textit{E. coli} consumes arabinose first and then xylose.
Figure 1.1. Carbon catabolite repression in *Escherichia coli* (Görke and Stülke, 2008)

Desai *et al.* showed that when both L-arabinose and D-xylose are present, AraC bound to arabinose substitutes XylR from the *xylR* promoters repressing expression of the genes required to transport and metabolize D-xylose (Desai and Rao, 2010) (Figure 1.2).

Apart from the regulatory proteins, small non-coding RNAs play significant roles in the sugar utilization pathways in bacteria (Görke and Vogel, 2008). The sRNA, SgrS is used by *E. coli* in response to stress induced by intracellular accumulation of D-glucose-6-phosphate. SgrR senses the build-up of D-glucose-6-phosphate and activates the expression of SgrS. SgrS then binds and degrades ptsG mRNA in presence of the RNA-binding protein Hfq (Vanderpool and Gottesman, 2004). Degradation of ptsG mRNA transcripts would lead to reduced transport and phosphorylation of D-glucose into the cells, decreasing the concentration of D-glucose-6-phosphate inside the cells. Also, another sRNA Spot 42 contributes to catabolite repression by repressing the catabolic and transport genes, thereby
introducing another layer of control in addition to the regulation of translation by CRP (Beisel and Storz, 2011).

In addition to regulation, modes of sugar transport can vary widely among the utilization pathways. Bacteria engage different energy driven transport systems to bring sugars inside the cells. Depending upon how they are linked to the energy source, the bacterial transport systems can be divided into primary and secondary transport systems. The primary transport systems are coupled to an energy generating mechanism such as ATP hydrolysis whereas the secondary transport systems are driven by an existing electrochemical gradient such as proton or sodium gradients. When there are no transporters available, then sugars can get into the cells via diffusion. While D-xylose, for example, uses a high affinity transporter XylF/G/H and a low affinity transporter XylE, L-rhamnose has only one transporter RhaT (Lam et al., 1980; Sofia et al., 1994; Vía et al., 1996). The high affinity transporter XylF/G/H is an ATP-binding cassette (ABC) transporter and is an example of primary transport system in \textit{E. coli}. The ABC transporters couple the energy released through ATP hydrolysis to move solutes into or out of the cells (Davidson and Chen, 2004). The low affinity transporter XylE, in contrast, is a proton symporter and is an example of secondary transport system in \textit{E. coli}. Another common mode of sugar transport in bacteria is the phosphotransferase system or PTS (Gosset, 2005; Kotrba et al., 2001). Unlike the other modes of transport, PTS transports sugar into the cells and simultaneously phosphorylates it. Considering the varying characteristics of natural sugar utilization pathways, restricting the study to only two sugar utilization pathways have severely limited our understanding of natural sugar utilization pathways.
1.4 Overview of the thesis
In the previous sections, the prevalence of individuality in bacteria and the importance of understanding individuality in bacterial sugar utilization have been discussed. Chapter 2 is a review article published in the journal *Chemical Engineering Science*. This review article mainly discussed the attributes of positive and negative feedback alone or in combination in various gene regulatory networks. Also, it elaborated the application of feedback in metabolic engineering, population control and biosensors and argues for the potential applications of feedback in synthetic circuits. Chapter 3 is a research article published in
Molecular Microbiology. In this article, the single-cell response of *E. coli* to eight different sugars was investigated using a combination of experimental and computational approaches. The response of *E. coli* to sugars was remarkably diverse, with negative feedback due to catabolism playing a crucial role. Also it discusses the effects of catabolism on system properties such as hysteresis. Chapter 4 is about engineering cells to convert sharp bimodal responses of natural sugar utilization pathways into linear, titratable responses. The motivation behind the work was to use the natural inducible sugar utilization pathways as ‘in-built’ titratable systems in non-model organisms. Using the L-arabinose utilization pathway as a model system, we showed that each modification came with a trade-off. The article was published in the journal *ACS Synthetic biology*. Chapter 5 is an extension of chapter 4 and shows how the inadvertent presence of inducer in the media could affect the response properties such as linearity and dynamic range. This is a manuscript in preparation. Finally, chapter 6 describes how this thesis can generate additional avenues of research.
References


Chapter 2
Understanding and Exploiting Feedback in Synthetic Biology

*This work is published in Chemical Engineering Science
Abstract

Synthetic biology employs traditional engineering concepts in the construction of cells and organisms. One of the most fundamental concepts is feedback, where the activity of a system is influenced by its output. Feedback can imbue the system with a range of desirable properties such as reducing the rise time or exhibiting an ultrasensitive response. Feedback is also commonly found in nature, further supporting the incorporation of feedback into synthetic biological systems. In this review, we discuss the common attributes of negative and positive feedback loops in gene regulatory networks, whether alone or in combination, and describe recent applications of feedback in metabolic engineering, population control, and the development of advanced biosensors. The examples principally come from synthetic systems in the bacterium *Escherichia coli* and in the budding yeast *Saccharomyces cerevisiae*, the two major workhorses of synthetic biology. Through this review, we argue that biological feedback represents a powerful yet underutilized tool that can advance the construction of biological systems.
2.1 Introduction

The field of synthetic biology aims to construct biological systems in order to better understand nature and to address pressing challenges facing society. However, the inherent complexity of biology impedes the construction of predictable and robust systems. Fortunately, concepts in traditional engineering disciplines offer approaches to reduce biological complexity and subsequently advance biological design.

One engineering discipline that has strongly influenced synthetic biology is process control. This discipline seeks to automate industrial processes in order to maintain system specifications with limited oversight. One prevalent tool in process control is feedback. Feedback serves multiple purposes, including driving the output of a system toward a desired setpoint, countering disturbances in system inputs, filtering measurement noise, and decoupling relationships between multiple inputs and multiple outputs. The physical link between system inputs and outputs is called a feedback loop. Process control principally focuses on negative feedback loops that drive the system output toward the setpoint. Positive feedback loops, which drive the system output away from the setpoint, are adopted less frequently (Horowitz and Hill, 1989).

Both negative and positive feedback loops can be found throughout the architecture of gene regulatory networks. Extensive studies, particularly in the field of systems biology, have revealed that these biological feedback loops shape the dynamics, variability, and steady-state response of the system. These influences in turn have been implicated in the adaptability and robustness of biological systems. The mechanisms of feedback vary widely and can be combined, resulting in overlaid negative and positive feedback loops with unique properties.

Despite the many benefits of feedback in biology and its prevalence in other engineering disciplines, synthetic biology has been slow to implement feedback in the design of biological systems. Arguably, the field is still grappling with how to construct large-scale systems that behave predictably and has not yet advanced to the point of including additional layers of control. As a result, recent work has centered on the construction of logic gates with either higher-order processing or the ability to interpret multiple signals (Rinaudo et al.,
Feedback has been integrated into a handful of synthetic biological systems, but often only in cases in which feedback was essential for the desired system behavior (e.g. genetic oscillators (Elowitz and Leibler, 2000; Atkinson et al., 2003; Fung et al., 2005; Stricker et al., 2008)). However, feedback offers a wealth of attributes that can generate more desirable behaviors and advance the robustness of biological systems.

In this review, we discuss the known properties of negative and positive feedback loops – both in isolation and in combination – in biological systems. Examples principally are drawn from bacteria and yeast, the current workhorses of synthetic biology. Many of the insights into feedback were drawn from synthetic systems, demonstrating how synthetic biology informs our understanding of nature. We then discuss recent applications of feedback loops in synthetic biology. Finally, we conclude by describing how feedback can be further implemented to advance current applications in the field.

2.2 Modes of Feedback

Feedback can occur at multiple steps of gene expression or through the interactions between organisms. Figure 2.1 illustrates representative mechanisms through which feedback can be introduced. We touch on many of these mechanisms in discussing the properties and application of feedback in biology. Note that even simpler examples of feedback often engage multiple mechanisms at one time.

2.2a Transcriptional regulation

In the first step of gene expression, messenger RNA is transcribed from the DNA of a gene. This step begins with transcription factors recruiting RNA polymerase to the gene’s promoter and ends with termination and release of the polymerase. Both steps offer opportunities for both negative and positive feedback, by controlling the accessibility or methylation of the DNA (Lim and Van Oudenaarden, 2007; Octavio et al., 2009), the availability of RNA polymerase or transcription factors, or transcriptional termination (Winkler et al., 2002; Lucks et al., 2011). The simplest form of transcriptional feedback, called auto-regulation,
involves a transcription factor regulating its own transcription. Auto-regulation is one of the most common transcriptional architectures found in bacteria (Shen-Orr et al., 2002) and has undergone the most extensive characterization out of all modes of feedback.

2.2b Post-transcriptional regulation
Following transcription, the messenger RNA is translated into protein. In microorganisms, this step can be regulated by modulating RNA stability and translation. In most cases, the responsible mechanisms involve the interaction of a trans-acting factor. This factor could be a protein, such as a ribonuclease or the RNA binding protein CsrA; an RNA, such as Hfq-binding small RNAs or synthetic riboregulators; or a small molecule, such as cofactors recognized by translational riboswitches (Waters and Storz, 2009). Many of the RNA-based mechanisms are still undergoing characterization and have not been quantitatively studied in the context of feedback.

2.2c Post-translational regulation
Once a protein is formed, feedback can be introduced by modulating the protein’s stability, localization, or activity. Stability can be modulated by attaching a degradation tag or altering protease activity, influencing protein levels. Next, localization can be modulated by targeting the protein to the cell membrane or to an organelle, affecting whether the protein can access its target. Finally, protein activity can be modulated through chemical modifications or reversible binding of small molecules, RNAs, or proteins ((Liu et al., 1997; Wassarman and Storz, 2000; Buchler and Cross, 2009; Hunsicker et al., 2009), altering the ability of the protein to carry out its normal functions. Modulating protein activity is the most common post-translational mechanism of feedback. One prominent example involves a small molecule reversibly binding to a transcription regulator, altering the ability of the regulator to bind DNA. Small-molecule responsive regulators, such as TetR, LacI, and AraC, are the most common components used in the construction of synthetic biological systems and are featured prominently in this review. Regardless of the exact mechanism, the protein inevitably influences its own activity or abundance. For instance, tetracycline binding to the
transcription repressor TetR de-represses the expression of the TetA transporter protein, which reduces the intracellular tetracycline concentration by pumping it out of the cell.

2.2d Cell-cell interactions

The final mode of feedback occurs between cells, influencing gene expression and/or cell growth. One common means of cell-cell interactions is through the secretion of diffusible factors. The secreted factors, either small molecules or peptides, interact with a surface receptor or an intracellular transcription regulator. In natural systems, one of these genes often is responsible for the synthesis of the signaling molecule, forming a positive feedback loop formed between all members of the population (Waters and Bassler, 2005). Induction and feedback only occur at high cell densities, granting cells a way to coordinate gene expression with their neighbors. This phenomenon is called quorum sensing (Ng and Bassler, 2009).

Quorum sensing is a standard tool in synthetic biology to elicit communication between engineered cells. The most common quorum sensing system, adapted from the marine bacterium Vibrio fischeri, combines the LuxI enzyme that synthesizes the N-acyl homoserine lactone (AHL) signaling molecule and the LuxR regulator that activates gene expression when bound to AHL. This system and other similar systems have been employed in a range of synthetic biology applications, including spatial patterning (Basu et al., 2005; Sohka et al., 2009), controlling the cell density of a bacterial culture (You et al., 2004), and engineering communication between bacteria and eukaryotes (Weber et al., 2007; Biliouris et al., 2012). As discussed later in this review, quorum sensing also offers a unique opportunity to introduce feedback across length scales well beyond that of a single cell.

Cell-cell interactions also can elicit feedback by modulating cellular health and survival. The interactions can be beneficial, such as complementary metabolic pathways, or harmful, such as predator-prey relationships. Many of these interactions can be readily engineered into a population of otherwise non-interacting organisms. Examples include linking quorum sensing to the activation or repression of a toxin (You et al., 2004; Balagaddé et al., 2008) and overproducing an essential metabolite that cannot be synthesized in other
The interacting cells form key components of feedback by ensuring the survival or death of their neighbors. This stands in contrast to the other modes of feedback that begin and end with gene expression.

**Figure 2.1** Modes of feedback in microorganisms. Feedback can be introduced at the different steps of gene expression (A-C) and through the interactions between cells (D). Red block lines designate negative regulation while green arrows designate positive regulation. (A) Transcriptional regulation, as shown for auto-repression. (B) Post-transcriptional regulation, as shown for an RNA-binding protein inhibiting its own translation. (C) Post-translational regulation, as shown for phosphorylation of a transcription repressor that inhibits DNA binding. (D,E) Cell-cell interactions, as shown for the synthesis and secretion of a diffusible molecule that activates the expression of the synthesis enzyme (D), and for a mutualistic interaction where one strain produces an essential metabolite not produced in the other strain (E).

### 2.3 Properties of Negative Feedback

Process control commonly employs negative feedback among other control strategies to maintain chemical processes. A feedback controller calculates the difference between the
measured output and a desired output of the system, the setpoint, such as product purity or temperature. The controller then adjusts the process to minimize this difference (proportional control). Aside from using the deviation from setpoint, the controller can take into account how long the difference has persisted (integral control) and how quickly the difference is changing (differential control). These different modes of feedback can be used depending on the required speed and stability by which the controller responds to setpoint deviations (Seborg et al., 2011).

Natural biological systems also implement negative feedback, but in ways that can differ from negative feedback in process control. First, in process control, negative feedback is often separate from the process itself through the actions and programming of an installed control system. In biological systems, the feedback is imbedded within the genetics and biochemistry of the process. Second, in process control, the principal goal of negative feedback is to maintain the system output within desired specifications. In biology, this goal can be more varied. Negative feedback alters the dynamics, steady-state behavior, and cell-cell variability of biological systems (Figure 2.2), where the benefit of each contribution likely depends on the cellular context (Thomas et al., 1995; Wall et al., 2003). Fortunately, we can exploit these attributes in the rational design of synthetic biological systems as discussed below. Note that all of the attributes we describe were observed for negative auto-regulation (Alon, 2007a), although other regulatory mechanisms have been shown to exhibit similar properties (Bashor et al., 2008).

2.3a Dynamics: Shorter response time

Microorganisms constantly face unpredictable and fluctuating environments. How quickly microorganisms adapt to new environmental conditions could mean the difference between life and death. Negative feedback offers one approach to influence the overall dynamics by shortening the rise time (Savageau, 1974). Rise time can be defined as the time required for the gene product to proceed halfway between an initial steady-state concentration and a final, elevated steady-state concentration. Rosenfeld and coworkers experimentally demonstrated the shorter rise time with a synthetic auto-repression circuit in Escherichia coli (Rosenfeld et
They tested two genetic circuits: one composed of the TetR transcription repressor controlling the expression of the GFP fluorescent reporter protein (direct regulation), and a second circuit composed of a TetR-GFP fusion that represses its own expression (auto-repression). Time course experiments following addition of the inducer, anhydrotetracycline (aTc), demonstrated that the rise time was up to five times smaller for auto-repression than for direct regulation. The reason for the increase in the rise time is the following: for promoters of equal strength, both GFP and the TetR-GFP fusion accumulate at the same rate. However, the TetR-GFP fusion begins repressing its own synthesis, lowering the final steady-state concentration. The result is an apparent acceleration toward the final steady-state. Accordingly, use of lower aTc concentrations further reduced the response time, because TetR still can repress its own expression. Note that the settling time, the time required for the gene product to proceed halfway between an initial steady-state concentration and a final, lowered steady-state concentration, is expected to be similar for auto-repression (e.g. through the removal of aTc) and for direct regulation. Despite this, the shorter response time afforded to negative feedback may be beneficial in the design of synthetic systems that must act rapidly, such as medical diagnostics.

2.3b Steady-state: Linearized response
The steady-state relationship between the input and output values of a negative feedback loop is just as important as the approach to steady state. This relationship can be illustrated for the bacterial transcription regulator TetR, which represses the expression of itself and the tetracycline exporter TetA. The antibiotic tetracycline binds TetR, relieving repression of both TetR expression and TetA expression. TetA then exports tetracycline, alleviating the inhibitory effect of this antibiotic on protein translation (Biliouris et al., 2011). While negative feedback should shorten the response time of the system to tetracycline, negative feedback also impacts the relationship between TetA levels and tetracycline concentrations. This relationship is important because it determines how well cells can negate the deleterious effects of tetracycline without investing too many resources in TetA production.
Nevozhay and coworkers investigated the steady-state attributes of negative feedback through the construction of two synthetic gene circuits in the model yeast *Saccharomyces cerevisiae* (Nevozhay et al., 2009). One circuit constitutively expressed TetR (direct regulation) while the other circuit auto-repressed TetR (auto-repression). Each circuit was tested by incubating the cells in varying concentrations of the tetracycline analog doxycycline and by measuring GFP under the control of a TetR-repressible promoter. They found that auto-repression extended the range of doxycycline concentrations that yield intermediate levels of GFP. In other words, auto-repression linearized the steady-state response to doxycycline. An accompanying mathematical model predicted this behavior, where negative feedback dampened the induction of the repressor. Madar and coworkers made similar observations for auto-repression of the AraC transcription regulator as part of L-arabinose utilization in *E. coli* (Madar et al., 2011). The benefit of linearization is that it allows the output to be finely tuned by varying the input signal. The ability to fine tune expression would be extremely useful in synthetic biology, such as when optimizing enzyme levels in an engineered metabolic pathway or when accurately predicting the concentration of a measured analyte with engineered biosensors.

### 2.3c Cell-cell variability: Noise reduction

Biological systems are inherently noisy (Balázsi et al., 2011). Noise, as measured by cell-cell variability in protein levels, arises from the small number of molecules involved in gene expression (intrinsic noise) and variation in the abundance and activity of the gene expression machinery (extrinsic noise). The result of such noise is large variations in protein levels between genetically identical cells in the same environment. These differences can interfere with the reliable processing of biological signals. In addition, these differences can introduce heterogeneity that promotes survival of a fraction of a cell population even in unpredictable and changing environments. Correspondingly, gene circuits have been shown to suppress or enhance noise, which will be critical for the future design of synthetic biological systems.
The impact of negative feedback on noise was first assessed experimentally by Becskei and Serrano (Becskei and Serrano, 2000). They evaluated a synthetic system in which a TetR-repressed promoter controlled the expression of a TetR-GFP fusion. Using fluorescence microscopy to measure GFP levels in individual cells, they found that the variability of GFP levels across the population was two to three times higher for direct regulation than for auto-repression. Nevozhay and coworkers further confirmed this finding in yeast (Nevozhay et al., 2009). The physical explanation is that auto-repression drives GFP levels to the same average value. When repressor levels are below this value, expression increases due to reduced feedback; when repressor levels are above this value, expression decreases due to enhanced feedback. Implementing negative feedback to reduce noise would be critical for any synthetic system that must faithfully convert an input signal into a defined output value. Reducing noise would be especially important when developing devices composed of only a few engineered cells, such as those for in vivo disease diagnostics in which population averaging cannot be applied.

One critical feature of negative feedback is the inherent delay between the altered input signal and the change in the system output (Seborg et al., 2011). Substantial delays in feedback, such as from the slow loss of a stable protein or from a cascade of regulators, can cause the system output to oscillate (Thomas, 1978; Bratsun et al., 2005). Stricker and coworkers experimentally demonstrated that even auto-repression, the simplest form of negative feedback, can exhibit sufficient delays in gene expression to cause oscillations in the measured output (Stricker et al., 2008). Maithreye and coworkers separately demonstrated that a delay in feedback increases instability of the system output, although systems with or without delayed feedback eventually converged on the same output value (Maithreye et al., 2008).
Figure 2.2 Contrasting characteristics of negative and positive feedback. These characteristics are also discussed in (Alon, 2007a). (A) Comparing direct regulation (gray), negative feedback (red), and positive feedback (green) according to (B) response dynamics, (C) input-output relationship at steady-state, and (D) cell-cell variability for an intermediate input value. In comparison to direct regulation, negative feedback and positive feedback exhibit opposite behaviors. The only exception is the settling time, where all three are predicted to exhibit similar response times. The indicated Hill coefficients assume no cooperativity in the feedback loop. The Hill coefficient and K in (C) are part of the Hill equation (Equation 1). Adapted from (Alon, 2007b).

The impact of delays begs the question: what is the propensity of feedback in an experimental biological system to exhibit oscillations and how much effort must be exerted to either avoid or promote this behavior? Further investigations will help clarify when
engineers need to be concerned about the potential for oscillatory behavior while introducing feedback into a biological system.

2.4 Properties of Positive Feedback
While process control inherently employs negative feedback, it tends to avoid positive feedback because of its destabilizing effect on the system output. The most common example in chemical processes is thermal runaway, where heat emitted by an exothermic reaction increases the reaction rate, releasing more heat. Thermal runaway poses a major safety threat in the chemical industry, and, when left uncontrolled, has led to numerous instances of catastrophic explosions and the release of chemical pollutants and toxins into the environment (Stoessel, 2008).

In contrast, positive feedback is widely employed in biology. For instance, bacteria employ positive feedback in sugar utilization, where the imported sugar induces the expression of sugar-specific transporters, leading to further sugar import. What distinguishes positive feedback in process control and in biological systems is that the latter always has an upper limit, preventing the biological equivalent of thermal runaway. Below we discuss the properties of positive feedback in biological systems in the same contexts as negative feedback: dynamics, steady-state behavior, and cell-cell variability (Alon, 2007a). Overall, the properties of positive feedback and negative feedback are opposites (Kaufman and Thomas, 2003), offering contrasting tools in the construction of synthetic biological systems (Figure 2.2).

2.4a Dynamics: Longer response time
In contrast to negative feedback, positive feedback exhibits a longer rise time in comparison to direct regulation (Savageau, 1974). The extended rise time was first demonstrated experimentally in E. coli using the transcription activator cI, which is involved in the lysis-lysogeny decision in lambda bacteriophage (Maeda and Sano, 2006). To generate positive feedback, Maeda and Sano tested a gene encoding a cI-GFP fusion under the control of a cI-activated promoter. Binding by the LacI repressor downstream of the promoter prevented
initial auto-activation, which could be relieved with the addition of the LacI inducer IPTG. For direct regulation, GFP was expressed from a LacI-repressed promoter. The rise time following the addition of IPTG was up to four times greater for positive feedback than for direct regulation. The explanation, as supported by mathematical modeling, was that positive feedback increased the steady-state output because more cI leads to greater transcription. Note that the settling times for direct regulation and auto-activation (e.g. removal of IPTG) are expected to be similar, paralleling deactivation of a negative feedback loop. A delayed response could be useful in the construction of synthetic systems that require a sustained input signal or a timed series of cellular events (Temme et al., 2008), although other genetic circuits can generate delays following circuit activation or deactivation (Mangan and Alon, 2003). Furthermore, positive feedback imbues a system with other critical behaviors as described below. The availability of other delay-generating circuits and the additional properties of positive feedback argue against the use of positive feedback in biological design with the sole intent of altering the response dynamics.

2.4b Steady-state: Ultrasensitivity

While negative feedback linearizes the response of a system to its input, positive feedback can give rise to an ultrasensitive response. Under this type of response, a small change in the input signal results in a much larger change in the output. This relationship is captured by the Hill coefficient (n), a constant in the following empirical equation relating the system input (I) and the system output (O):

\[ O = O_{\text{min}} + \left( O_{\text{max}} - O_{\text{min}} \right) \frac{I^n}{K^n + I^n}, \]

where \( K \) is the input value yielding an output halfway between the maximal output (\( O_{\text{max}} \)) and the minimal output (\( O_{\text{min}} \)). This equation can be fit to experimental data in order to estimate a value of the Hill coefficient. For a system lacking any feedback or cooperativity, the Hill coefficient is one. For Hill coefficient values greater than one, the response curve becomes steeper, indicating that the response has become more sensitive.
One of the best-studied examples of ultrasensitivity in positive feedback is the lactose utilization pathway. This pathway exhibits positive feedback when induced with the lactose analog IPTG. IPTG relieves repression by the transcription repressor LacI, which induces expression of the lactose transporter LacY and leads to the further import of IPTG. Jensen and coworkers first demonstrated a sharp response of the activity of a LacI-repressed promoter to IPTG, which became shallower in the absence of lacY (Jensen et al., 1993). A Hill coefficient of 4.5 later was measured for this same system, compared to 2.6 in the absence of lacY (Kuhlman et al., 2007). To put this into perspective, for a Hill coefficient of one, the IPTG concentration must increase by a factor of 81 to go from 10% to 90% of the maximal output. For a Hill coefficient of 4.5, the IPTG concentration must increase by a factor of 2.7 to achieve the same relative change in output, underscoring how positive feedback can render the relationship between the system input and output more switch-like. The reason for this behavior is that positive feedback amplifies the input signal, further increasing the system output.

Positive feedback may be beneficial for a range of synthetic biology applications requiring switch-like responses, such as the development of digital cellular devices or the threshold detection of disease biomarkers. It is worth noting that other circuit configurations and modes of regulation can produce ultrasensitivity, such as protein allostery (Koshland et al., 1966), regulatory cascades (Huang and Ferrell, 1996; Hooshangi et al., 2005), or sequestration or stoichiometric action of a regulator (Levine et al., 2007; Buchler and Cross, 2009). However, none of these other configurations have a propensity for bistability as described next for positive feedback.

2.4c Cell-cell variability: Increased noise and bistability

Positive feedback increases cell-cell variability in a similar fashion to other circuit configurations exhibiting ultrasensitivity (Hooshangi et al., 2005; Mehta et al., 2008). The increased variability emerges from noise in the input leading to large changes in the output. However, in the presence of strong positive feedback, a biological system can become bistable (Guespin-Michel and Kaufman, 2001). Bistability was observed experimentally over
fifty years ago in the lactose utilization pathway (Novick and Weiner, 1957). Intermediate concentrations of the lactose analog thio-methylgalactoside (TMG) induced the pathway either fully or negligibly, a phenomenon that has been called an ‘all-or-none’ response. The ‘all-or-none’ response emerges from the import of TMG inducing the expression of the transporter LacY, which drives further leads TMG import and LacY expression. This cycle of import and induction continues until LacY levels are maximized. Similar bistable behavior was observed in a natural MAPK signaling cascade that induces the expression of the MAPK components (Ferrell and Machleder, 1998) and in synthetic transcriptional auto-activation systems in E. coli (Isaacs et al., 2003) and in budding yeast (Becskei et al., 2001; Ajo-Franklin et al., 2007).

One common feature of bistable systems with rate exception is hysteresis (Guidi and Goldbeter, 1997). Hysteresis represents a phenomenon where the input-output relationship is influenced by the history of the system. In the case of the lactose utilization pathway, pre-incubating the cells with TMG lowered the TMG concentration associated with the transition between induced and uninduced states (Novick and Weiner, 1957; Ozbudak et al., 2004). The basis of hysteresis is that positive feedback maintains the induced state even at lower inducer levels. Only through stochastic fluctuations do the cells switch states (Choi et al., 2008). When positive feedback is sufficiently strong, the cell can be ‘locked’ into the induced state even upon complete loss of the input signal. Members of the Silver laboratory exploited this feature to generate synthetic ‘memory’ devices that capture a transient cellular state such as DNA damage (Ajo-Franklin et al., 2007; Burrill and Silver, 2011).

Mathematical modeling has shown that non-linearity in the positive feedback loop – whether through non-linear kinetics or regulator cooperativity – is essential for bistability and hysteresis (Keller, 1995; Ozbudak et al., 2004). However, there are exceptions to this rule. To and Maheshri demonstrated in yeast that auto-activation by a non-cooperative regulator can generate a steady-state bimodal response as long as the regulator is unstable and expressed stochastically (To and Maheshri, 2010). However, hysteresis was not explored, leaving the unanswered question of whether their system exhibits bistability. Separately, Tan and coworkers demonstrated in bacteria that auto-activation of a non-cooperative regulator can
generate bistability and hysteresis as long as overexpression of the regulator slows cellular growth (Tan et al., 2009). Note that the decreased cellular growth introduced an additional positive feedback loop, which improved the robustness of bistability as discussed in section 5.2.

Bistability is arguably the most advantageous attribute of positive feedback for synthetic biology because it ensures that biological systems either are fully induced or uninduced—a digital readout of an input signal. The ability to establish one of two fixed states would benefit a wide range of synthetic biology applications, including programmed cell differentiation, analog-to-digital conversions for cellular computing, and generation of synthetic microbial consortia from a single starter strain.

2.5 Properties of Combined Negative and Positive Feedback

While negative feedback and positive feedback offer contrasting attributes, combining these forms of feedback can augment these attributes or introduce entirely new ones (Demongeot et al., 2000). We next discuss the ramifications and potential applications of combining feedback loops in biology. Our focus is on the most salient features of each combination. Figure 2.3 illustrates features of representative loops. Overall, each combination can offer improved properties over a single positive or negative feedback loop, although the improvements come at the cost of additional system components and the potential for additional emergent behaviors.

2.5a Multi-negative feedback loops

In the two seminal examples of synthetic genetic circuits, negative regulatory events were combined in series into a single feedback loop. In the first example, Gardner and coworkers constructed a double-negative feedback loop in *E. coli* (Gardner et al., 2000). This ‘toggle switch’ was comprised of the transcription repressors LacI and TetR, where LacI repressed the expression of TetR and TetR repressed the expression of LacI. Because a double-negative feedback loop is equivalent to a single positive feedback loop, the toggle switch similarly exhibited ultrasensitivity, bistability, and hysteresis. However, unlike positive feedback, the
toggle switch could fix each stable state – in this case either high LacI levels or high TetR levels. Furthermore, a pulse of inducer, either IPTG or aTc, was sufficient to drive the system.

**Figure 2.3** Combining negative and positive feedback. The presence of multiple regulatory interactions imparts enhanced or even unique properties. (A) Double-negative and triple-negative feedback loops. A double-negative feedback loop (left) functions similarly to a single positive feedback loop, only the double-negative loop ‘locks’ the induced state and the uninduced state. Transient exposure to the appropriate inducer drives the system from one state to another. A triple-negative feedback loop (right) generates oscillations, although the oscillations are inconsistent. (B) Layered positive feedback loops. These loops exhibit the same general properties of ultrasensitivity and bistability as a single positive feedback loop. However, both properties in the layered loops are more robust to changes in system parameters such as feedback strength or expression strength. (C) Nested negative and positive feedback loops. Depending on the exact configuration and system parameters, the system can exhibit a range of behaviors including stable oscillations (left) and excitability (right).
into the opposite state. The ability to drive the system irreversibly into either state offers a clear advantage over auto-activation, where only the induced state is reinforced.

In the second example, Elowitz and Leibler constructed and tested a triple-negative feedback loop in *E. coli* that has been dubbed the ‘repressilator’ (Elowitz and Leibler, 2000). This loop was comprised of three transcription regulators, TetR, cI, and LacI, where TetR repressed the expression of cI, cI repressed the expression of LacI, and LacI repressed the expression of TetR. This loop resembles a single-negative feedback loop with a delay in feedback. Accordingly, mathematical modeling predicted that this circuit would exhibit oscillations under a broad parameter range (Elowitz and Leibler, 2000). While the circuit did exhibit oscillations, the oscillations were short-lived with ranging amplitudes and periods. This unsatisfactory performance was attributed to noise in gene expression. The properties of the ‘repressilator’ argue against the use of triple-negative feedback loops to generate oscillations, especially in comparison to other oscillatory circuits that combine negative and positive feedback (see section 2.5c).

### 2.5b Layered positive feedback loops

Unlike multi-negative feedback loops, any number of positive regulatory steps results in positive feedback. The properties described above for a single positive feedback loop thus apply to multiple positive feedback loops. The notable difference is that the presence of multiple positive feedback loops can improve the robustness of these properties. Shah and Sarkar comprehensively investigated the properties of multiple feedback loops by modeling the behavior of all possible interactions between two or three regulators (Shah and Sarkar, 2011). In particular, they evaluated the extent to which each possible circuit exhibits ultrasensitivity and bistability across a range of parameter values, which can be described as the robustness of the behavior. Both enzyme and transcription regulators were included in the analysis to account for differences in the regulatory properties. Modeling revealed that the presence of multiple positive feedback loops increased the robustness in both ultrasensitivity and bistability. Furthermore, the most robust circuits were hybrid circuits of both enzyme regulators and transcription regulators. Part of the rationale for the improved robustness was
that enzyme regulators such as kinases and phosphatases can exhibit zero-order ultrasensitivity (Goldbeter and Koshland, 1981) that contributes to both ultrasensitivity and bistability.

It is worth noting that regulatory cascades, regulator sequestration, and stoichiometric action by a regulator can also generate ultrasensitive responses. When coupled with positive feedback, these mechanisms have been shown to enhance the robustness of bistability (Ferrell and Machleder, 1998; Bashor et al., 2008; Levine and Hwa, 2008). Chen and Arkin most recently demonstrated this effect in *E. coli* using a repressor (the anti-sigma factor *rsiW* from *Bacillus subtilis*) that sequesters an auto-activated regulator (the sigma factor *sigW* from the same bacterium expressed from a SigW-activated promoter) (Chen and Arkin, 2012). Titrating the anti-sigma factor tuned the input threshold that separates the induced and uninduced states, offering a means to readily tailor the quantitative properties of this system. These characteristics would be beneficial in synthetic biology for the construction of bistable systems without exhaustive component optimization.

2.5c Nested negative and positive feedback loops

Combining both negative and positive feedback intertwines otherwise opposing attributes, potentially creating altogether unique and flexible responses. One response is oscillations, where competition between negative and positive feedback forces the system to oscillate between induced and uninduced states. The oscillations are more stable than those associated with a multi-negative feedback loop (e.g. the ‘repressilator’), arguing for one utility of nested negative and positive feedbacks. Atkinson and coworkers first demonstrated this behavior in *E. coli* by combining positive feedback (auto-activation by the transcription regulator NtrC) and negative feedback (NtrC-activated expression of LacI that represses the expression of NtrC) (Atkinson et al., 2003). The resulting synthetic circuit exhibited oscillations, although these oscillations dampened within a few cycles. Fung and coworkers demonstrated that similar oscillations could be generated by combining transcriptional and metabolic regulation in *E. coli* (Fung et al., 2005). In their system, dubbed the ‘metabolator,’ the conversion between two metabolic pools was controlled by two enzymes that were regulated either
positively or negatively by one of the metabolites. This system also exhibited oscillations, although the amplitude and persistence of the oscillations varied.

Advancing on these previous efforts, Stricker and coworkers generated stable oscillations in *E. coli* using the transcription activator AraC and the transcription repressor LacI (Stricker et al., 2008). Specifically, they designed a system with one positive feedback loop (AraC auto-activation) and two negative feedback loops (LacI auto-repression and AraC-activated expression of LacI that represses the AraC expression). Their design showed remarkable consistency in the amplitude and period of the oscillations, where the period could be tuned by changing temperature, media conditions, or the concentration of the inducers L-arabinose and IPTG. The improved persistence and uniformity of the oscillations may be attributed to the additional feedback loops built into the circuit. As an extension of this design, members of the same research group constructed a similar circuit using quorum sensing (Danino et al., 2010). This system also contained both positive feedback (activation of the AHL synthase LuxI by AHL) and negative feedback (activation of the AHL-degrading enzyme AiiA by AHL) in a similar configuration to the previous design. The new design synchronized oscillations across the entire population, resulting in macroscopic oscillations in otherwise microscopic cells.

Coupling positive and negative feedback loops can also lead to excitability, a distinct response not observed for positive or negative feedback loops in isolation. In this response, the system undergoes stochastic and transient induction. Stochastic fluctuations in the mediator of positive feedback drive induction until negative feedback returns the system to the uninduced state. Excitable behavior was observed directly in DNA uptake in *B. subtilis* during nutrient starvation (Süel et al., 2006). Two feedback loops control ComK, the master regulator of the uptake machinery. ComK undergoes auto-activation (positive feedback) and also represses ComS, which otherwise inhibits the degradation of ComK (negative feedback). In roughly 4% of the population of starved cells, stochastic fluctuations in ComK led to an increase in ComK levels driven by auto-activation. Eventually, reduction in ComS levels caused the increased degradation of ComK, returning ComK levels to the original uninduced state. Through the construction of synthetic circuits, Çağatay and coworkers revealed that
other regulatory configurations of the same components exhibit excitability, although the natural circuit exhibited the greatest and most variable duration of induction. The distinct induction characteristics of the natural circuit allowed for greater variability in the uptake of DNA between cells, perceivably balancing the risk and reward of random DNA uptake (Cağatay et al., 2009). Synthetic biology could benefit from stochastic and transient induction in the development of altruistic biological systems (Lee et al., 2010) or of systems requiring a consistent subpopulation regardless of growth conditions.

The combination of negative and positive feedback can lead to a range of behaviors depending on the exact circuit configuration and system parameters. Tian and coworkers performed modeling to assess how the relative strength of negative and positive feedback impact the resulting response (Tian et al., 2009). They found that the relative strength of negative and positive feedback is the predominant determinant of the observed response. Bistability emerges when positive feedback dominates, whereas a graded, unimodal response emerges when negative feedback dominates – exactly paralleling what was observed for each type of feedback in isolation. When the strengths of feedback are balanced, the system can exhibit excitability or oscillations. The varied responses demonstrate that coupling negative and positive feedback offers a flexible motif for system design and may be the best approach for generating stable oscillations or excitability.

### 2.6 Current Applications of Feedback in Synthetic Biology

Clearly, negative feedback and positive feedback offer diverse advantages in the construction of synthetic biological systems. However, there are only a few examples of feedback being applied in synthetic biology to address scientific or technological challenges. We discuss representative examples below, which fall into three categories: metabolic control, biosensor design, and population control (Figure 2.4).

#### 2.6a Metabolic control

Metabolic engineering has overwhelmingly relied on static control for the expression of heterologous pathways (Tyo et al., 2007). Static control tunes the levels of each enzyme to
maximize product yields, although the levels can be changed only through further genetic manipulations. In contrast, cells are dynamical systems that undergo constant perturbations even in a well-controlled environment, similar to chemical processes. Perturbations such as variability in enzyme levels or in metabolite levels can have a detrimental impact on the overall yield of the final product. Typically, these perturbations result in the diversion of too much metabolic flux away from cellular growth or in the build-up of toxic intermediates in the culture medium. One solution is the use of metabolic feedback to dynamically regulate enzyme levels. Farmer and Liao first demonstrated the benefits of metabolic feedback in *E. coli* in order to regulate the microbial conversion of glucose to the carotenoid lycopene (Farmer and Liao, 2000). They employed the transcription activator NtrC, which is phosphorylated by the metabolic intermediate acetyl phosphate. Because acetyl phosphate levels increase as excess flux from glucose catabolism is directed toward acetate production, NtrC activity could be employed to divert excess metabolic flux from acetate production to product synthesis. To create this link, Farmer and Liao placed genes involved in lycopene synthesis under an NtrC-activated promoter. This genetic manipulation created a negative feedback loop that reduced lycopene production when too much metabolic flux was diverted away from cell growth and increased lycopene production when metabolic flux was being wasted through acetate production. Introduction of this loop improved cell growth and increased the yield of lycopene by at least ten-fold, offering a generalized approach to balance growth demands and product synthesis in a wide range of metabolic engineering applications.

Zhang and coworkers adopted a different approach to introduce dynamic metabolic control into microbial chemical synthesis, specifically for the conversion of glucose to fatty acid ethyl esters in *E. coli* (Zhang et al., 2012). Rather than balancing cell growth and product synthesis, they controlled two convergent branches of the fatty acid ethyl ester synthesis pathway with fatty acid, a key pathway intermediate. The genes encoding the pathway branches were placed under the control of the transcription repressor FadR, which is deactivated when bound to fatty acids or fatty acyl-CoA. The result was a negative feedback loop that increased fatty acid conversion when the concentration of fatty acid was elevated.
This loop increased the yield of fatty acid ethyl esters by three-fold and improved the stability of pathway genes. It is worth noting that one of the inhibitors of FadR, fatty acyl-CoA, is synthetized from fatty acid, inadvertently introducing a positive feedback loop that may generate bistability or oscillations. Observing these behaviors would require single-cell analyses, which are rarely performed in metabolic engineering. Overall, these studies demonstrate the utility of negative feedback in the dynamic control of heterologous metabolic pathways. Because metabolite-responsive transcription regulators and riboswitches are available in nature and can be engineered, dynamic control could be applied broadly in metabolic engineering (Tang et al., 2008; Carothers et al., 2011; Michener et al., 2012). Applying these regulators for metabolic control will advance metabolic engineering by consistently improving product yields and making microbial chemical synthesis a more competitive alternative to traditional routes of industrial chemical synthesis.

2.6b Biosensor design

Feedback was an essential feature of the next category of applications. Building on previous work from the Hasty group, Prindle and coworkers developed a biosensor that modulates its oscillatory frequency in response to a sensed molecule (Prindle et al., 2012). The underlying genetic circuit coupled negative and positive feedback through two diffusible signaling molecules: AHL and hydrogen peroxide. AHL induced the expression of the AHL synthase LuxI (positive feedback) and the AHL-degrading enzyme AiiA (negative feedback). Hydrogen peroxide activated the expression of NADH dehydrogenase II responsible for hydrogen peroxide production (positive feedback). A hybrid promoter controlled by both AHL and hydrogen peroxide regulated the circuit output GFP. *E. coli* cells harboring the engineered circuit were grown in a microfluidic device that collected individual micro-colonies or ‘biopixels.’ A permeable wall and gas-filled cavity separated adjacent micro-colonies. AHL coordinated oscillations within the micro-colony, while hydrogen peroxide coordinated oscillations between adjacent micro-colonies by diffusing through the cavity. This setup synchronized oscillations between cells in the same micro-colony as well as
adjacent colonies, leading to synchronous oscillations over millimeter-length scales – over 1,000 times the size of an individual bacterial cell.

Prindle and coworkers expanded this engineered circuit to sense arsenic. An additional copy of the luxI gene was placed under the control of an arsenic-responsive promoter, which is activated when soluble arsenite binds the transcription regulator ArsR. The inducible expression of additional LuxI modulated the period of the oscillations from 68 minutes in the absence of arsenite to 80 minutes in the presence of 1 μM arsenite. Importantly, the period closely correlated with intermediate arsenite concentrations, offering a macroscopic, frequency-based readout of toxin concentrations. The range of periods admittedly was limited, although insights from the group’s previous work (Stricker et al., 2008) coupled with riboswitches or other post-transcriptional regulatory mechanisms (Liang et al., 2011) could expand the observed range, thereby improving the accuracy and sensitivity of frequency-based biosensors.

2.6c Population control
In the final category of applications, feedback was applied to engineer interactions between cells that impact growth and viability. This category can be split into two types: antagonistic interactions, which restrain population growth, and mutualistic interactions, which promote population growth. In the simplest example of antagonistic interactions, You and coworkers engineered a negative feedback loop in E. coli that induces cell death at high cell densities (You et al., 2004). The loop connected quorum sensing to the expression of the toxic protein CcdB. At low cell densities, the expression of the ccdB gene was low and growth rate of the cells was unperturbed. At high cell densities, the expression of the ccdB gene was activated, preventing the further accumulation of cells. The plateau in cell density exhibited damped oscillations in line with a mathematical model and could be tuned by modulating the stability of AHL.

Balagaddé and coworkers extended this work to construct a synthetic predator-prey system composed of two engineered strains of E. coli (Balagaddé et al., 2008). This system combined two orthogonal quorum sensing systems and either induction or repression of the
The engineered ‘predator’ *E. coli* cells secreted one AHL molecule and repressed the expression of *ccdB* in the presence of a different AHL signaling molecule. The engineered ‘prey’ *E. coli* cells secreted the second AHL molecule and activated the expression of *ccdB* in the presence of the first AHL signaling molecule. This configuration generated a negative feedback loop composed of two cells where the prey increased the growth of the predator, while the predator reduced the growth of the prey. The synthetic circuit replicated common behaviors associated with natural predator-prey systems, including coexistence, oscillations, and extinction, that depended on the experimental conditions. Aside from providing insights into natural systems, the two antagonistic circuits conferred control over the density of an isogenic or mixed population. The major limitation is that the antagonistic interactions exert selective pressure on the cells, potentially driving the loss of the *ccdB* gene in the above examples (Balagaddé et al., 2005). This limitation would pose challenges for the long-term maintenance of a population density.

Mutualistic interactions offer a distinct approach to control cellular populations. These interactions promote the growth of all participating members of the population, which is a common attribute of microbial consortia found in nature (Wintemute and Silver, 2010). In one example of an engineered mutualistic system, Shou and coworkers engineered two yeast strains to form a positive feedback loop through cross-feeding (Shou et al., 2007). One strain was engineered to overproduce the nucleobase adenine and to not produce the essential amino acid lysine, while the other strain was engineered to overproduce lysine and to not produce adenine. Engineering each strain to produce one metabolite and not the other
Figure 2.4 Current and future applications of feedback in synthetic biology. (A) Metabolic control. Feedback can be used to control enzyme expression and activity based on the concentration of pathway intermediates or metabolic flux toward cell growth and homeostasis. The illustrated example controls enzyme expression for two branches of a metabolic pathway. Control limits both excessive enzyme expression and the toxic buildup of the intermediate. Adapted from (Zhang et al., 2012). (B) Advanced biosensors with accurate and measurable outputs. This example couples negative and positive feedback for stable oscillations. The sensed molecule controls the expression of one component of the loop, thereby modulating the frequency of oscillations. (C) Population control. Feedback can be engineered into the interactions between different organisms. The interaction can help restrain the relative number of cells even if one organism grows faster than the other. Adapted from (Balagaddé et al., 2008). (D) Robust and rapid logic devices. Feedback can be introduced to accelerate processing by logic gates or to generate a more switch-like response. A three-input AND gate serves as an example. Adapted from (Moon et al., 2012).
interlinked their growth. Following inoculation of the two strains into a single culture, the
cell density of both strains showed large initial swings. Interrogation of the large swings
revealed that each strain had to undergo lysis in order to release the essential metabolite, an
intriguing feature of the system. Following the swings in cell density, the systems showed
persistent stability even after large dilutions, which the authors attributed to the selection of
beneficial mutations in either strain.

In a more recent example of engineered mutualistic interactions, Kerner and
coworkers generated two strains of *E. coli* that excrete a metabolite required by the opposing
strain (Kerner et al., 2012). In contrast to the circuit developed by Shou and coworkers,
excretion of the metabolites did not require cell killing and was under inducible control.
Growth of the strains with different concentrations of the exogenous inducers tuned the
growth rate of co-culture and the ratio of the two strains. The resulting relationship was
highly non-linear and did not lend to a simple mathematical model. We attribute this
relationship to the competing benefits (utilizing the essential metabolite) and costs (gene
over-expression and metabolite shedding) of the mutualistic interactions, potentially
imparting nested positive and negative feedback loops.

These two circuits demonstrate that positive feedback, through mutualistic
interactions can stabilize the ratios of participating strains. These interactions are expected to
be more stable than antagonistic interactions by promoting rather than restraining growth.
The limitation to this approach is that participating strains will continue growing until the
resources in the medium are depleted. Future work within this application could focus on
combining both mutualistic and antagonistic interactions to control population density and
composition and to achieve novel population dynamics afforded to coupling positive and
negative feedback loops. The design of engineered communities undoubtedly will become a
central thrust of synthetic biology with widespread applications in metabolic engineering,
human health, and biomanufacturing.
2.7 Future directions

The above examples illustrate how feedback has been integrated into the design of biological systems. However, feedback has been absent in one of the largest thrusts of synthetic biology: the design of logic devices. These devices are being pursued with the goal of constructing increasingly large genetic circuits that perform more complex logic operations. Recent examples have marked major milestones in engineering biology, including the development of circuits that count (Friedland et al., 2009), circuits that process up to four different input signals (Moon et al., 2012), and circuits that detect edges separating regions in the light and in the dark (Tabor et al., 2009). These circuits can be described as first-pass ‘prototypes,’ where more advanced versions perceivably would integrate control strategies, fail-safes, and other more complex operations. However, the prototypes tended to require extensive optimization to function properly even under defined laboratory conditions. Some of the optimization may be attributed to typical complications such as poor expression, which can be relieved through recently developed approaches (Salis et al., 2009; Lou et al., 2012; Qi et al., 2012). However, the need for optimization could be relaxed through the integration of feedback in the initial design. For instance, logic devices could be constructed with multiple positive feedback loops to introduce robust and tunable bistability or with negative feedback loops to reduce cell-cell variability in the sensory components (Figure 2.4D). Integration of feedback may even reduce the optimization time, leading to the faster development of a functional system. To facilitate the widespread use of feedback, a library of portable feedback modules could be developed that exhibit a range of feedback properties and can readily be inserted into existing genetic circuits (Nistala et al., 2010). These modules could be as simple as auto-repression or as complex as layered positive feedback. Modules also could offer a range of feedback properties through genetic modifications or use of allosteric regulators that can be tuned exogenously (Kerner et al., 2012). Finally, these modules could be extended to feature other modes of feedback, such as allosteric enzymes used in the optimization of metabolic pathway design. The resulting collection of modules could be broadly utilized for the construction of biological systems with relaxed design requirements.
Another potential hurdle to integrating feedback is the predominant use of transcriptional regulation in logic devices. Promoters have a limited capacity for regulator binding sites, restricting how many inputs a single promoter can receive. This restriction could be alleviated through the further development of post-transcriptional regulators such as regulatory RNAs. Ongoing work in the design and construction of synthetic regulatory RNAs should contribute significantly to the construction of large-scale circuits and the accompanying ability to introduce feedback (Win and Smolke, 2008; Lucks et al., 2011). For this idea to be fully realized, we must better understand how regulatory RNAs shape the properties of feedback. Recent work has highlighted how regulatory RNAs and transcription regulators can display different regulatory properties, even in the context of genetic circuits (Levine and Hwa, 2008; Mehta et al., 2008; Beisel and Storz, 2011). Further efforts to understand the properties of both synthetic and natural RNA-based genetic circuits will lay the foundation to engineer feedback with diverse mechanisms of regulation.

2.8 Conclusions

Biological feedback offers numerous advantages for synthetic biology that extend beyond the typical use of feedback in process control. Many of these advantages were elucidated through the characterization of synthetic biological systems. Even though synthetic biology has lent critical insight into the attributes of feedback, the field has been slow to adopt these attributes. As constructed systems become larger and more complex, feedback will become an essential feature that must be integrated even in the prototype stage. Doing so will facilitate future efforts in biological design, potentially matching or even surpassing the prevalence of feedback in nature.

Acknowledgements

We thank K. Biliouris, S. Carrell, A. Jermusyk, and G. Reeves for critical reading of the manuscript. T.A. and C.L.B. currently are supported through start-up funds provided by North Carolina State University.
References


Chapter 3

Bacterial sugar utilization gives rise to distinct single-cell behaviors

*This work has been published in Molecular Microbiology. Contributions to this work:
  Konstantinos Biliouris, Yiannis Kaznessis
Abstract

Inducible utilization pathways reflect widespread microbial strategies to uptake and consume sugars from the environment. Despite their broad importance and extensive characterization, little is known how these pathways naturally respond to their inducing sugar in individual cells. Here, we performed single-cell analyses to probe the behavior of representative pathways in the model bacterium *Escherichia coli*. We observed diverse single-cell behaviors, including uniform responses (D-lactose, D-galactose, N-acetylglucosamine, N-acetylneuraminic acid), “all-or-none” responses (D-xylose, L-rhamnose), and complex combinations thereof (L-arabinose, D-gluconate). Mathematical modeling and probing of genetically modified pathways revealed that the simple framework underlying these pathways—inducible transport and inducible catabolism—could give rise to most of these behaviors. Sugar catabolism was also an important feature, as disruption of catabolism eliminated tunable induction as well as enhanced memory of previous conditions. For instance, disruption of catabolism in pathways that respond to endogenously synthesized sugars led to full pathway induction even in the absence of exogenous sugar. Our findings demonstrate the remarkable flexibility of this simple biological framework, with direct implications for environmental adaptation and the engineering of synthetic utilization pathways as titratable expression systems and for metabolic engineering.
3.1 Introduction

All forms of life must obtain sources of carbon and energy from the environment in order to survive and replicate. Within microorganisms, a prevalent strategy is the use of inducible utilization pathways that uptake and consume exogenous sugars. These pathways encode transporters and enzymes that are expressed only in the presence of the sugar (Figure 3.1A). This framework can be viewed as a cost-saving measure to conserve resources when the sugar is absent (Dekel and Alon, 2005). However, this framework also imparts positive and negative feedback (Figure 3.1B) (Krishna et al., 2007). Positive feedback emerges from the sugar inducing the expression of transporters, which in turn import additional sugar molecules into the cell. Similarly, negative feedback emerges from the sugar inducing the expression of catabolic enzymes, which in turn shunt the intracellular sugar into metabolism.

Coupling positive and negative feedback has been shown to impart unique properties not ascribed to either mode of feedback alone (Thomas, 1978; Tian et al., 2009; Shah and Sarkar, 2011; Afroz and Beisel, 2013). For instance, natural and synthetic gene regulatory networks that feature positive and negative feedback have been reported to exhibit ranging behaviors (Brandman and Meyer, 2008), such as stable oscillations within synthetic transcriptional networks (Atkinson et al., 2003; Stricker et al., 2008; Tigges et al., 2009; Prindle et al., 2012; Biliouris et al., 2012), excitability within the uptake of foreign DNA (Süel et al., 2006), and bistability with limited memory within the transcriptional response to D-galactose in yeast (Acar et al., 2005; Avendaño et al., 2013).

Although inducible utilization pathways inherently couple positive and negative feedback, these pathways have been assumed to only exhibit uniform behaviors (Ninfa and Mayo, 2004; Smits et al., 2006; Krishna et al., 2007). This assertion can be traced to (i) the predominant use of bulk characterization techniques (Kuhlman et al., 2007; Kaplan et al., 2008), (ii) single-cell analyses of the D-lactose and L-arabinose utilization pathways deprived of sugar catabolism (Novick and Weiner, 1957; Cohn and Horibata, 1959; Siegele and Hu, 1997; Khlebnikov et al., 2001; Morgan-Kiss et al., 2002; Ozbudak et al., 2004; Megerle et al., 2008), and (iii) limited mathematical modeling without any ensuing experimental validation (Krishna et al., 2007). The prevailing notion is that, in the absence of
catabolism, these pathways exhibit “all-or-none” responses in which pathways are maximally induced (“all”) or negligibly induced (“none”) in single cells; in the presence of catabolism,

**Figure 3.1** Inducible sugar utilization. (A) Inducible utilization pathways are commonly composed of a set of transporters and catabolic enzymes whose expression is induced in the presence of the recognized sugar. (B) The pathways harbor a combination of positive (green plus) and negative (red dash) feedback. Positive feedback emerges from the sugar inducing the expression of the transporters, leading to additional import of the sugar. Negative feedback emerges from the sugar inducing the expression of the enzymes, leading to breakdown of the sugar.

the “all-or-none” response is lost, yielding a uniform response among all cells in the population. However, the only single-cell analyses of a native utilization pathway (L-arabinose) revealed a bimodal response even in the presence of sugar catabolism (Khlebnikov et al., 2001; Morgan-Kiss et al., 2002). Therefore, what remains unclear is how different utilization pathways respond to their inducing sugar in individual cells and the role of sugar catabolism in the response.

Here, we investigated the single-cell behaviors of representative inducible utilization pathways within the model bacterium *E. coli*. Surprisingly, the assessed pathways exhibited
varying single-cell behaviors, including uniform responses, “all-or-none” responses, and varying combinations thereof. Within the diversity of responses, we found that the simple coupling of inducible transport and inducible catabolism could give rise to most of these behaviors. Furthermore, sugar catabolism was an important contributor to each behavior, where loss of catabolism eliminated tunable induction and enhanced memory of previous conditions. These findings demonstrate the remarkable flexibility of this simple biological framework and suggest that individual pathway behaviors can be tailored—either through evolution or through engineering—to sugars present in the environment.

3.2 Results

3.2a Varying single-cell behaviors in sugar utilization in E. coli

We focused on inducible utilization pathways in E. coli K-12 substrain MG1655 because this bacterium possesses at least 16 pathways containing both inducible transport and inducible catabolism and virtually all of the pathway components have been identified and well characterized (Table S3.1). In particular, we concentrated on eight pathways (lactose, D-galactose, N-acetylglucosamine, N-acetylneuraminic acid, D-xylose, L-rhamnose, L-arabinose, and D-gluconate) that have undergone extensive characterization. To monitor the single-cell response in each pathway, we placed the green fluorescent protein gene (gfp) under the control of a pathway promoter on a low-copy plasmid (Zaslaver et al., 2006). The use of this plasmid parallels many other studies of gene regulation in single cells (Khlebnikov et al., 2001; Morgan-Kiss et al., 2002; Ozbudak et al., 2004) and reflects a trade-off between a strong fluorescence signal and titration of the pathway components (Del Vecchio et al., 2008). MG1655 cells transformed with each plasmid were grown in M9 minimal medium with 0.4% glycerol (M9 glycerol) with varying concentrations of the sugar before being subjected to flow cytometry analysis (Figure S3.1A). The cells were grown in exponential phase for 20 hours to a low final density (ABS$_{600}$ ~ 0.002) to help ensure that the fluorescence values approached a stable distribution (Figure S3.1B) and to minimize depletion of the sugar in the medium (Ozbudak et al., 2004; Avendaño et al., 2013). Dot plots of the resulting distributions were generated from the fluorescence histograms similar to
previous work (Pelet et al., 2011). The dot plots communicate unimodal or bimodal distributions as well as the fluorescence and relative size of each sub-population (Figure S3.1A).

Remarkably, the assessed utilization pathways exhibited ranging responses (Figure 3.2). The D-lactose, D-galactose, N-acetylglucosamine, and N-acetylneuraminic acid utilization pathways exhibited graded responses wherein all cells exhibited uniform fluorescence levels at all sugar concentrations. The D-xylose and L-rhamnose utilization pathways exhibited “all-or-none” responses wherein the sugar concentration only affected the relative fraction of cells in the fully induced or uninduced states (Ozbudak et al., 2004). The L-arabinose utilization pathway exhibited a bimodal response at low sugar concentrations but then a graded response at higher sugar concentrations once the entire population was induced. Finally, the D-gluconate utilization pathway exhibited a graded response at lower sugar concentrations only to bifurcate into induced and uninduced sub-populations at higher sugar concentrations. The responses appeared to be independent of the specific promoter selected, as another promoter for D-xylose (P$_{araF}$) and three other pathway promoters for L-arabinose (P$_{araE}$, P$_{araF}$, P$_{araJ}$) exhibited the same qualitative response (compare Figure 3.2 and Figure S3.2A). The responses also did not correlate with the doubling time of each strain in the presence or absence of sugar (Table S3.2), downplaying the potential contribution of growth rate. Finally, the responses did not appear to be influenced by the copy number of the reporter plasmid based on single integrants of the P$_{araB}$ reporter, although lower fluorescence levels complicated interpretation of the response (Figure S3.2B).
Figure 3.2 Varying single-cell responses to different sugars in E. coli. MG1655 cells harboring a transcriptional reporter plasmid were grown in M9 glycerol with varying concentrations of the indicated sugar for 20 hours in exponential phase to a low cell density and then subjected to flow cytometry analysis. The promoter used to drive GFP expression is specified. Each dot represents the mean fluorescence and the relative number of cells in the induced (black) and uninduced (white) subpopulations (see Figure S3.1A). Only one dot is shown for unimodal distributions. The diameter of each dot scales with the fraction of cells in that population. Each dot plot is representative of at least three biological replicates conducted on separate days.

3.2b A simple model predicts varying behaviors from the underlying framework

The observed behaviors may be due to attributes unique to each pathway, such as feedback through central metabolism (e.g. carbon catabolite repression), the number of transport systems and enzymatic routes, or the presence of additional regulatory mechanisms. However, another possibility is that the shared framework of inducible transport and inducible catabolism underlies many of these behaviors. To explore the potential contribution of this framework, we developed a simple deterministic model of a generalized utilization pathway (see Supplementary Information, Table S3.3). This pathway consisted of one
transporter and one enzyme whose expression is induced by a sensory activator in the presence of intracellular sugar (Figure 3.3A). To simplify the analysis, the model (i) assumed similar dilution rates for the transporter, enzyme, and sugar through cell division and (ii) neglected sugar-dependent differences in growth rate. Within the model, the strength of positive and negative feedback were captured by the maximal activities of the fully expressed transporter ($\alpha$) and enzyme ($\gamma$), respectively. Unlike previous models of sugar utilization (Carrier and Keasling, 1999; Ozbudak et al., 2004; Krishna et al., 2007; Santillán and Mackey, 2008; Stamatakis and Mantzaris, 2009; Savageau, 2011; Yildirim, 2012), this model is not focused on any single pathway, accounts for inducible sugar catabolism, and does not overly simplify the mechanisms of sugar import and catabolism.

The model predicted graded as well as bistable responses that principally depended on the strengths of inducible transport (positive feedback) and inducible catabolism (negative feedback). As shown in the phase diagram in Figure 3.3B, graded responses were predicted for weak transport activity (low $\alpha$) as well as for strong catabolic activity (high $\gamma$), whereas bistable responses were predicted for strong transport activity and weak catabolic activity (high $\alpha$ and low $\gamma$). These behaviors follow from positive feedback driving bistability and negative feedback countering bistability as predicted for transcriptional networks (Tian et al., 2009). Excitability or oscillations were also predicted for unequal dilution rates of the transporter, enzyme, and intracellular sugar (Figure S3.3). Similar behaviors were also observed for a sensory repressor (Figure S3.4, Supplementary Information) or for two transporters (Afroz et al., 2014). These predictions suggest that the combined influence of inducible import and catabolism of an exogenous sugar can give rise to different emergent behaviors. Inducible import and catabolism of an exogenous sugar can give rise to different emergent behaviors. The model also predicted unique features of the response curves—the relationships between extracellular sugar and enzyme levels—within the phase diagram. First, minimal transport activity was necessary to achieve a reasonable dynamic range, defined here as the ratio of enzyme levels with saturating or no extracellular sugar (Figure 3.3B, dashed blue line). This minimal value always fell within the bistable region for lower catabolic activities ($\gamma < 0.048$), suggesting that sugar catabolism principally determines
whether an inducible pathway will exhibit graded or bistable behavior. Second, we observed that modulating the catabolic activity had other effects on the response (Figure 3.3C). In particular, increasing the catabolic activity within the bistable regime (i) expanded the tunability of the induced state and (ii) reduced the width of the bifurcation region, representing the range of extracellular sugar concentrations in which two stable enzymes levels were possible. In contrast, altering transport activity had a minor impact on the inducible import and catabolism of an exogenous sugar can give rise to different emergent behaviors. The model also predicted unique features of the response curves—the relationships between extracellular sugar and enzyme levels—within the phase diagram. First, minimal transport activity was necessary to achieve a reasonable dynamic range, defined here as the ratio of enzyme levels with saturating or no extracellular sugar (Figure 3.3B, dashed blue line). This minimal value always fell within the bistable region for lower catabolic activities (γ < 0.048), suggesting that sugar catabolism principally determines whether an inducible pathway will exhibit graded or bistable behavior. Second, we observed that modulating the catabolic activity had other effects on the response (Figure 3.3C). In particular, increasing the catabolic activity within the bistable regime (i) expanded the tunability of the induced state and (ii) reduced the width of the bifurcation region, representing the range of extracellular sugar concentrations in which two stable enzymes levels were possible. In contrast, altering transport activity had a minor impact on the
**Figure 3.3** A simple model predicts varying responses from coupling inducible transport and inducible catabolism. (A) Major reactions considered by the deterministic model. Extracellular sugar (S₀) is imported into the cell by the transporter (T). The intracellular sugar (S) can reversibly bind to the sensory activator or can undergo irreversible degradation by the enzyme (E). The activator bound to the sugar (R) upregulates the expression of the transporter and the enzyme. See Supplementary Information for more details on the model. (B) Phase diagram based on the maximal activities of the fully expressed transporter (α) and enzyme (γ). See Table S3.3 for values of fixed parameters (with ν = 0). Monostable regions represent parameter values in which the model predicts one stable enzyme level for any extracellular sugar concentration at steady-state. Bistable regions represent parameter values in which the model predicts two stable and one unstable enzyme levels for at least one extracellular sugar concentration at steady-state. The dashed blue line represents parameters that achieve 95% of the maximal dynamic range, where the dynamic range is defined as the ratio of enzyme levels for saturating extracellular sugar and no sugar. Proceeding to the right of this line further increases the dynamic range. (C) Bifurcation curves representing the steady-state relationship between enzyme levels and extracellular sugar concentration for a given set of parameter values. Different curves are shown whereby the only altered parameter value is the maximal enzyme activity (γ). Dashed red lines indicate steady-state enzyme levels that are unstable. (D) Bifurcation curves whereby the only altered parameter value is the maximal transporter activity (α). See Supplementary Information for more details about the model.
inducible import and catabolism of an exogenous sugar can give rise to different emergent behaviors. The model also predicted unique features of the response curves—the relationships between extracellular sugar and enzyme levels—within the phase diagram. First, minimal transport activity was necessary to achieve a reasonable dynamic range, defined here as the ratio of enzyme levels with saturating or no extracellular sugar (Figure 3.3B, dashed blue line). This minimal value always fell within the bistable region for lower catabolic activities ($\gamma < 0.048$), suggesting that sugar catabolism principally determines whether an inducible pathway will exhibit graded or bistable behavior. Second, we observed that modulating the catabolic activity had other effects on the response (Figure 3.3C). In particular, increasing the catabolic activity within the bistable regime (i) expanded the tunability of the induced state and (ii) reduced the width of the bifurcation region, representing the range of extracellular sugar concentrations in which two stable enzymes levels were possible. In contrast, altering transport activity had a minor impact on the tunability of the induced state as well as the width of the bifurcation region (Figure 3.3D). These predictions mirror most of the observed behaviors (Figure 3.2), supporting the contribution of positive and negative feedback.

3.2c Sugar catabolism allows tunable induction

To probe model predictions, we experimentally investigated how sugar catabolism contributes to the observed response. We focused on all but the D-lactose and D-gluconate utilization pathways (Figure 3.4): the D-lactose pathway was previously interrogated in the absence of catabolism using the non-hydrolyzable lactose analog TMG (Novick and Weiner, 1957; Cohn and Horibata, 1959; Ozbudak et al., 2004), whereas the D-gluconate pathway contains redundant catabolic steps through the two D-gluconate kinases GntK and IdnK (Istúriz et al., 1986; Vivas et al., 1994). For each evaluated pathway, we disrupted sugar catabolism by deleting the catabolic genes, which eliminated the ability of $E. coli$ to grow on each sugar (data not shown). This disruption generally reduced growth in the presence of the sugar (Table S3.2), as expected when inducing pathway expression without any fitness benefit. We then generated response curves for the resulting catabolism-deficient (ΔE) strains under the same growth conditions as that of the wild type (WT) strains (Figure 3.2). Note that
the induction time of 20 hours was sufficient to achieve a stable distribution as determined for L-arabinose (Figure S3.1B).

Flow cytometry analysis of the catabolism-deficient strains revealed two general responses lacking tunable induction (Figure 3.4): (i) an “all-or-none” response in which cells were either negligibly or fully induced (N-acetyleneuraminic acid, D-xylose, L-rhamnose, L-arabinose), and (ii) full induction independent of the sugar concentration (D-galactose, N-acetylglucosamine). The “all-or-none” response paralleled the simulated response curves in the absence of sugar catabolism ($\gamma = 0$) and previous single-cell studies of pathways lacking catabolism (Novick and Weiner, 1957; Siegele and Hu, 1997; Ozbudak et al., 2004). This

Figure 3.4 Single-cell responses in the absence of sugar catabolism. The specific genes deleted from each pathway and the promoter driving GFP expression are indicated. Note that the lactose utilization pathway was not evaluated because its behavior was previously probed in the absence of sugar catabolism through the use of the non-hydrolyzable inducer TMG (Ozbudak et al., 2004). The D-gluconate utilization pathway was also not evaluated because of redundancy in D-gluconate catabolism. See Figure 3.2 for the experimental growth conditions and an explanation of the dot plots. All dot plots are representative of at least three biological replicates conducted on separate days.
response also lacked tunability of the induced state, in line with setting $\gamma = 0$ at different points within the predicted phase diagram (Figure 3.3). Note that the loss of tunability for L-arabinose cannot be attributed to self-catabolite repression (Katz and Englesberg, 1971), as the presence of 2 mM cyclic AMP in the medium did not alter the qualitative response to L-arabinose (Figure S3.2C). Furthermore, excision of the resistance cassette used to delete the catabolic operons had no impact on the response (Figure S3.5). We thus attribute these behaviors to inducible transport driving bistability based on the loss of catabolism.

The D-galactose and N-acetylglucosamine utilization pathways lacking catabolism exhibited full induction even in the absence of exogenous sugar (Figure 3.4). A distinguishing feature of both pathways is that *E. coli* can synthesize each inducing sugar (Wu and Kalckar, 1966; Wu and Wu, 1971; Vogler *et al.*, 1989). In particular, N-acetylglucosamine can be synthesized through the sequential activities of L-glutamine:D-fructose-6-phosphate aminotransferase (GlmS) and N-acetylglucosamine deactylase (NagA), whereas D-galactose is an inadvertent bi-product of the Leloir pathway. To explore the impact of endogenous biosynthesis, we modified our mathematical model to allow for constitutive production of intracellular sugar (Figure 3.5A, Supplementary Information). The model predicted that sugar biosynthesis elevated expression levels in the absence of exogenous sugar and converted a bistable response into a graded response (Figure 3.5B). Both predictions emerge from the synthesized sugar inducing transporter expression even in the absence of exogenous sugar, thereby breaking positive feedback. The model also predicted that disrupting sugar catabolism elevated expression levels in the absence of exogenous sugar, which was magnified with increasing catabolic activity (Figure 3.5C). In this case, catabolism restrains build-up of the endogenous sugar, which would otherwise drive full induction of the pathway. These predictions closely match the experimental observations for D-galactose and N-acetylglucosamine (Figure 3.4), although leaky promoter activity may also contribute to the basal expression. Overall, catabolism appears to be a critical component of utilization pathways that respond to sugars synthesized by the cell.
3.2d Sugar catabolism reduces the extent of hysteresis

Another intriguing model prediction was that sugar catabolism reduces the extent of hysteresis—or memory of previous conditions—by counteracting bistability. In the case of inducible utilization pathways, hysteresis would reflect initially induced cells being more sensitive to the sugar than initially uninduced cells (Novick and Weiner, 1957). To experimentally test this prediction, we focused on the D-xylose and L-arabinose utilization pathways because they exhibited a clear bimodal response in the absence of sugar catabolism and were fully induced at high sugar concentrations. For each pathway, we included either no sugar or sufficient sugar to induce the entire cellular population in the overnight cultures of WT and ΔE cells. The cells were then washed to remove residual sugar, back diluted into varying sugar concentrations, and grown for 20 hours prior to flow cytometry analysis (Figure 3.6). To quantify the extent of hysteresis, we calculated the ratio of concentrations to achieve ~50% induced cells with or without pre-incubation of the sugar.

Paralleling model predictions, we found that deleting the catabolic genes substantially increased the extent of hysteresis for both pathways (Figure 3.6). In particular, deleting the catabolic genes increased the extent of hysteresis by a factor of over 100 for D-xylose (concentration ratios of 9.9 ± 3.0 and 1,770 ± 570 for the WT and ΔxylAB strains, respectively; average ± S.E.M., n = 3) and by a factor of over 4 for L-arabinose (concentration ratios of 1.0 ± 0.1 and 4.1 ± 0.6 for the WT and ΔaraBAD strains, respectively, average ± S.E.M., n = 3).
Figure 3.5 Impact of endogenous biosynthesis of the sugar on the predicted response. (A) The model was modified to allow for constitutive biosynthesis of the intracellular sugar (ν, yellow arrow). (B) Bifurcation curves for the predicted enzyme levels (E) in the presence of sugar catabolism (γ = 0.1) for different rates of endogenous sugar production. See Figure 3.2 for more information about the bifurcation curves. (C) Predicted ratio of enzyme levels without (γ = 0) or with (γ > 0) sugar catabolism in the absence of exogenous sugar (S₀ = 0). The ratio peaks at intermediate rates of endogenous sugar biosynthesis and then begins to decline as enzyme levels approach full induction (E = 1) even in the presence of sugar catabolism. All plots represent α = 0.1 and all other parameter values are reported in Table S3.3. See Supplementary Information for more details about the model.

The extent of hysteresis between these pathways was quite distinct, particularly in the absence of sugar catabolism (concentration ratios of 1,770 for D-xylose and 4.1 for L-arabinose). Interestingly, this difference cannot be readily attributed to the strength of positive feedback (α), which was predicted to have little effect on
Figure 3.6 Extent of hysteresis for the D-xylose and L-arabinose utilization pathways with intact or disabled sugar catabolism. Overnight cultures of wild type cells (WT) or cells lacking the catabolic genes (ΔxylAB or ΔaraBAD) were incubated with or without sufficient sugar to induce the entire population (1 mM of D-xylose or 100 μM of L-arabinose for WT, 0.1 mM of D-xylose for ΔxylAB, 10 μM of L-arabinose for ΔaraBAD). The cultures were then washed and grown in M9 glycerol with the indicated concentration of sugar for 20 hours. See Figure 3.2 for an explanation of the dot plots. The gray regions represent the extent of hysteresis based on the extrapolated sugar concentration required to induce ~50% of the population. All dot plots are representative of at least three biological replicates conducted on separate days.

the width of the bifurcation region (Figure 3.3D). Instead, sensitivity analysis identified the Hill coefficient for transporter expression (n1) as the largest contributor. This would suggest...
that apparent cooperativity in the expression of pathway transporters, such as through DNA looping (Kuhlman et al., 2007), accounts for the extent of hysteresis between pathways (Figure S3.6).

As an alternative means to evaluate hysteresis, we sorted bimodal populations of WT and ΔaraBAD cells into induced and uninduced sub-populations and monitored each sub-population in the same medium over time monitored by flow cytometry analysis (Figure S3.7). The WT sub-populations converged on similar fluorescence distributions, while the ΔaraBAD sub-populations maintained fixed fluorescence distributions despite active growth of the cultures (Figure S3.7). We thus conclude that a distinct contribution of sugar catabolism is reducing memory of previous conditions.

3.3 Discussion

We found that inducible utilization pathways exhibit diverse single-cell responses. This finding directly challenges the repeated assertion that all natural utilization pathways exhibit uniform behaviors (Ninfa and Mayo, 2004; Smits et al., 2006; Krishna et al., 2007) and underscores the need for single-cell analyses when studying inducible sugar utilization in the microbial world (Lubelska et al., 2006; Johnsen et al., 2009; Horák, 2013). Our findings also highlight the limitations of drawing conclusions about natural pathways from studies that neglect sugar catabolism. For instance, we found that the lactose utilization pathway naturally exhibits a graded response, yet this pathway is widely associated with an “all-or-none” response that emerges only in the absence of catabolism (Novick and Weiner, 1957; Ozbudak et al., 2004).

Our model predictions suggest that the combination of inducible transport and inducible catabolism in itself can give rise to diverse emergent behaviors. One ramification is that pathways may be tailored to conditions in which the inducing sugar is naturally found. For instance, some pathways may exhibit bimodal responses to sugars whose concentrations fluctuate erratically (Acar et al., 2008; Bennett et al., 2008), whereas other pathways may exhibit graded responses to sugars that are present more consistently. Alternatively, the
bimodal responses may allow coordinated sugar utilization while conserving resources (Kovárová-Kovar and Egli, 1998) or may promote heterogeneity as part of sugar-dependent processes such as virulence (Muñoz-Elías and McKinney, 2006; Rojo, 2010), biofilm formation (Sadykov et al., 2011; Chai et al., 2012), or the uptake of foreign DNA (Lo Scrudato and Blokesch, 2012). The complexity of the responses may be even further enhanced in the presence of multiple sugars because of carbon catabolite repression (Kaplan et al., 2008; Görke and Stülke, 2008).

(Kaplan et al., 2008; Görke and Stülke, 2008), cross-regulation between pathways (Desai and Rao, 2010), or promiscuity of some transport systems (Zhu and Lin, 1986; Kornberg and Lourenco, 2006). As one example, such responses would impact commensal *E. coli* as it consumes N-acetylneuraminic acid, N-acetylgulcosamine, D-gluconate, and other sugars during gut colonization (Chang et al., 2004).

Our deterministic modeling considered a simplified pathway of one transporter and one enzyme to draw general insights into inducible sugar utilization. However, each natural pathway is likely influenced by additional regulatory factors, such as auto-repression by the regulator (Hahn and Schleif, 1983; Weickert and Adhya, 1993; Egan and Schleif, 1993), multiple transport systems (Yildirim, 2012), signaling through downstream metabolic products (Katz and Englesberg, 1971), or gene repression by regulatory RNAs (Beisel and Storz, 2011; Papenfort et al., 2013). Furthermore, differential growth on the sugar (Table S3.2) can lead to faster or slower dilution of the cell components (Tan et al., 2009; Klumpp et al., 2009). Many of these factors would impart positive or negative feedback that may influence the single-cell behaviors. For instance, we speculate that the bimodal and non-monotonic response to D-gluconate (Figure 3.2) could emerge in part from (i) D-gluconate lowering cAMP levels (Epstein et al., 1975), which would downregulate pathway expression, and (ii) the overlapping metabolic routes involved in D-gluconate catabolism (Wu and Wu, 1971; Vogler et al., 1989). Beyond the framework considered in this work, other known frameworks of inducible utilization pathways exist, such as pathways induced by metabolic intermediates (e.g. maltose, L-fucose; see Table S3.1). These alternative frameworks may
result in disparate behaviors that warrant further investigation and may expand the diversity of single-cell responses associated with inducible sugar utilization.

Our insights are expected to inform the engineering of utilization pathways. Sugar utilization represents the first step in most endeavors within metabolic engineering (Alper and Stephanopoulos, 2009) and environmental bioremediation (Ramos et al., 2011). Depending on the abundance and identity of the sugar, only a fraction of the population may be actively importing and consuming the sugar. Any existing or imported utilization pathway would require interrogation at the single-cell level and potential manipulation to ensure a uniform response—for instance, by strengthening negative feedback or by disrupting positive feedback (Khlebnikov et al., 2001; Morgan-Kiss et al., 2002; Afroz et al., 2014). Similar issues may be encountered when co-opting natural utilization pathways as titratable expression systems (Guzman et al., 1995; Cardona et al., 2006). Finally, the combination of inducible transport/synthesis and export/degradation offers a simple framework for the construction of synthetic gene networks. The existing toolkit in synthetic biology is principally composed of gene regulatory devices such as transcription factors or regulatory RNAs. Therefore, the development of synthetic inducible pathways would greatly expand this toolkit and offers means to generate tunable or memory-based devices.

3.4 Experimental procedures

Bacterial strains and plasmids. All of the strains were derivatives of E. coli K-12 substrain MG1655 (obtained from S. Gottesman, National Cancer Institute). All strains, plasmids, and oligonucleotides used in this work are reported in Tables S3.4, S3.5, and S3.6, respectively. To generate the deletion strains, the chloramphenicol resistance cassette encoding the cat gene was amplified by PCR from pKD3 using primers with gene-specific homology regions at the 5’ ends (Datsenko and Wanner, 2000). The resulting linear PCR products were recombined into NM500 harboring the mini-λ cassette (Court et al., 2003) in order to generate NM500 ΔaraBAD::cat, NM500 ΔxylAB::cat, NM500 ΔrhaBADM::cat, NM500 ΔgalKM::cat, NM500 ΔnagB::cat, and NM500 ΔnanA::cat. Each resistance cassette was then transferred to MG1655 by P1 transduction. The resistance cassette in MG1655
ΔnagB::cat, MG1655 ΔnanA::cat, MG1655 ΔaraBAD::cat, and MG1655 ΔxylAB::cat was excised using pCP20 (Cherepanov and Wackernagel, 1995), thereby generating MG1655 ΔnagB, MG1655 ΔnanA, MG1655 ΔaraBAD -cat, and MG1655 ΔxylAB -cat respectively. Successful recombinants and transductants were verified by PCR and by sequencing.

All of the reporter plasmids were based on the pUA66 plasmid encoding the gfpmut2 variant of the green fluorescent protein gene downstream of a multi-cloning site (Zaslaver et al., 2006). The plasmid also encodes a kanamycin resistance cassette and the low-copy sc101 origin-of-replication. Promoter regions from the E. coli genome are inserted into the multicloning site, resulting in a fluorescent readout of transcriptional activity. Most of the plasmids were purchased through OpenBiosystems from the E. coli promoter collection originally developed by the Alon laboratory (Zaslaver et al., 2006). The rest of the plasmids (pUA66ParaE, pUA66ParaF, pUA66Pxy1A, pUA66Pxy1F) were unavailable in the collection. To generate pUA66Pxy1F, the promoter region was amplified from the MG1655 chromosome and cloned into the BamHI/XhoI restriction sites in pUA66. To generate pUA66ParaE, pUA66ParaF and pUA66xylA, the promoter regions were also amplified from the MG1655 chromosome by PCR. For these promotors, the primers were designed with 5’ ends that overlapped with the pUA66 plasmid. The backbone of pUA66 was amplified by PCR in two fragments (primers pUA66.fwd1/pUA66.rev1 and pUA66.fwd2/pUA66.rev2) from pUA66 DNA linearized with XhoI. The amplified PCR products were assembled using the Gibson method (Gibson, 2011). All cloned plasmids were verified by colony PCR and by sequencing.

To integrate the $P_{araB}$-gfp construct into the E. coli MG1655 genome, we amplified the reporter construct from pUA66ParaB by PCR and cloned the construct into pCAH63 (Haldimann and Wanner, 2001) in place of uidAf by the Gibson method. The resulting pCAH63-ParaB plasmid was then integrated into the E. coli genome as described previously with slight modifications (Haldimann and Wanner, 2001). Briefly, the plasmid was transformed at 30°C into MG1655 cells carrying the helper plasmid pINT-ts. The transformed cells were recovered at 37°C for one hour and then at 42°C for 30 minutes before plating on LB agar containing chloramphenicol to select for integrants at 37°C. Colonies
were then restreaked on LB agar at 42°C to clear the helper plasmid. Individual colonies were then replicate plated on LB agar containing chloramphenicol or ampicillin. Colonies that were resistant to chloramphenicol and sensitive to ampicillin were screened by PCR to identify single integrants (Haldimann and Wanner, 2001).

**Growth conditions and media.** All strains were streaked from freezer stocks onto LB plates and single colonies were inoculated into 2 ml of LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride). Each reporter plasmid was transformed into the indicated strain by electroporation followed by preparation of freezer stocks. For the flow cytometry experiments, freezer stocks of cells with the reporter plasmid were streaked to isolation on LB agar (LB media with 1.5% agar) supplemented with 50 μg/ml of kanamycin. Individual colonies were then inoculated into M9 minimal medium (1X M9 salts (Difco), 2 mM magnesium sulfate, 100 μM calcium chloride, 0.001% thiamine hydrochloride) with 0.4% glycerol and 50 μg/ml of kanamycin (M9 glycerol) and shaken at 250 RPM and 37°C. The cultures were grown to a final turbidity of ABS$_{600}$ ~1.0 as measured on a Nanodrop 2000c spectrophotometer (Thermo Scientific). These cultures were back diluted into 2 ml of M9 glycerol with varying concentrations of the indicated sugar and grown for 20 hours at 250 RPM and 37°C prior to flow cytometry analysis. Back dilutions (~ 1:10$^6$ for 20-hour induction experiments) were made such that the cell densities prior to flow cytometry analysis were ABS$_{600}$ ~0.002. Based on the measured doubling times (Table S3.2) and back dilutions for each strain, the cultures appeared to be in exponential phase growth for the vast majority of the incubation time. For the time-course experiments in Figure S3.1A, overnight cultures were back diluted to achieve a final turbidity of ABS$_{600}$ ~0.002 for each indicated incubation time. To measure hysteresis, the indicated concentration of sugar was included in the overnight cultures. The overnight cultures then were pelleted by centrifugation and resuspended in M9 glycerol two times prior to back dilution for the flow cytometry experiments.
Flow cytometry and generation of dot plots. Fluorescence measurements were performed on a BD Accuri C6 flow cytometer equipped with a CFlow plate sampler, a 488 nm laser, and a 530 ± 15 nm bandpass filter. Cultures were directly run on the cytometer without any dilutions or media changes. Events reflecting cells were gated based on forward scatter (FSC-H) and side scatter (SSC-H) using lower cutoffs of 11,500 A.U. and 500 A.U., respectively. The specific gate was set using DRAQ5 red dye to ensure that only cells and not debris appeared within the gate. At least 2,000 gated events were collected for each measurement based on the low cell density and the location of the gate. The green fluorescence of gated events was measured using the FL1 channel and recorded as FL1-H. The fluorescence distributions were fit with unimodal or bimodal normal distributions implementing the MATLAB function *fit*. The mean and area of each distribution were converted into dots. The mean determined the vertical location of the dot, whereas the area determined the size of the dot. The diameter of each dot directly scaled with the area under each normal distribution. Figure S3.1A illustrates the process of generating dot plots from the events recorded on the flow cytometer.

Cell sorting. Cells were grown for 20 hours in M9 minimal medium with 0.4% glycerol containing the indicated concentration of L-arabinose to give roughly 50% induced cells and ABS$_{600}$ ~0.002. Cells were then sorted into uninduced and induced populations on a MoFlo cell sorter (Beckman Coulter) with 1X PBS as the sheath fluid. Importantly, FACSFlow, the standard sheath fluid for Beckman Coulter, was not used because it arrested growth of *E. coli* unless extensively diluted. At least 1,000,000 cells in each population were collected. Sorted cells were back diluted different amounts into medium with the designated concentration of L-arabinose. Each back-diluted culture was grown until reaching an ABS$_{600}$ of ~0.002 and then subjected to flow cytometry analysis.

Mathematical modeling. Details on the deterministic model can be found in Supplemental Information, including its derivation and parameter values. Briefly, the model comprises three differential equations that capture sugar import and catabolism and the inducible
expression of the transporter and the enzyme. All simulations were conducted in Matlab using the Euler method and the arc-length continuation method.

Acknowledgements
We are grateful to M. Luo, G. Storz, and K. Volzing for critical reading of the manuscript. This work was supported by start-up funds and an external mentoring mini-grant from the NC State ADVANCE D3 program provided to C.L.B., a Doctoral Dissertation Fellowship awarded to K.B., and grants from the National Institutes of Health (GM086865), the National Science Foundation (CBET-0425882, CBET-0644792), the Minnesota Supercomputing Institute, and the National Computational Science Alliance (TG-MCA04N033) to Y.K.

The authors declare no conflicts of interest.
References


Chapter 3

Bacterial sugar utilization gives rise to distinct single-cell behaviors
(Supplementary Information)

*This work has been published in *Molecular Microbiology*. Contributions to this work:
Konstantinos Biliouris, Yiannis Kaznessis
Description of the mathematical model

We developed a simple deterministic model composed of four equations to describe the behavior of a generalized inducible sugar utilization pathway. The model accounts for the regulator (R), the intracellular sugar (S), one transporter (T), and one enzyme (E). The regulator can function either as an activator (e.g. AraC) or as a repressor (e.g. LacI). In the case where the regulator functions as an activator, binding of sugar to the regulator generates an active complex (R) that upregulates the expression of the enzyme and the transporter. The transporter imports additional sugar into the cell while the enzyme irreversibly degrades the intracellular sugar. The equations shown below describe the amount of R (eq. 1), S (eq. 2), T, (eq. 3), and E (eq. 4). The model further assumes a Hill function captures the relationship between active regulator and the expression of the enzyme and the transporter. By varying the Hill coefficient (n₁, n₂), this function can indirectly capture DNA looping and other sources of apparent cooperativity (Huo et al., 1988; Kuhlman et al., 2007) as well as auto-repression of the regulator common to most utilization pathways that linearizes the relationship (Madar et al., 2011). A separate feature common to many pathways is the presence of multiple transport systems. While quantitative differences were predicted for systems with one or two transport systems (Yildirim, 2012), we observed no qualitative differences in the predicted behavior when a second transporter was added to the model (Afroz et al., 2014).

\[
\begin{align*}
\text{R} &= \frac{S \text{active regulator}}{k_0 + S} \\
\frac{dS}{dt} &= v_{\text{endogenous production}} + \frac{S_0 T \text{active transport of sugar}}{k_1 + S_0} - \frac{S E \text{catabolism of sugar}}{k_2 + S} \quad \text{dilution of sugar} \\
\frac{dT}{dt} &= \frac{b_T \text{basal expression of transporter}}{k_3 + R^n} + \frac{R^n \text{induction of transporter by sugar}}{k_4 + R^n} \quad \text{dilution of transporter} \\
\frac{dE}{dt} &= \frac{b_E \text{basal expression of enzyme}}{k_5 + R^n} + \frac{R^n \text{induction of enzyme by sugar}}{k_6 + R^n} \quad \text{dilution of enzyme}
\end{align*}
\]
Each parameter in the model is described below.

**Table S 3.1 Parameters of the model**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description (Units)</th>
<th>Symbol</th>
<th>Description (Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Concentration of active regulator (M)</td>
<td>k₄</td>
<td>Half-maximum regulator concentration (M)</td>
</tr>
<tr>
<td>Rₜ</td>
<td>Concentration of total regulator (M)</td>
<td>bₜ</td>
<td>Basal expression of transporter (M/s)</td>
</tr>
<tr>
<td>S</td>
<td>Intracellular sugar concentration (M)</td>
<td>bₑ</td>
<td>Basal expression of enzyme (M/s)</td>
</tr>
<tr>
<td>S₀</td>
<td>Extracellular sugar concentration(M)</td>
<td>α₁</td>
<td>Maximum rate of sugar import (1/s)</td>
</tr>
<tr>
<td>t</td>
<td>Time (s)</td>
<td>α₂</td>
<td>Maximum rate of sugar catabolism (1/s)</td>
</tr>
<tr>
<td>T</td>
<td>Transporter concentration (M)</td>
<td>α₃</td>
<td>Maximum rate of transporter production (M/s)</td>
</tr>
<tr>
<td>E</td>
<td>Enzyme concentration (M)</td>
<td>α₄</td>
<td>Maximum rate of enzyme production (M/s)</td>
</tr>
<tr>
<td>k₀</td>
<td>Half maximum regulator concentration (M)</td>
<td>dᵢ</td>
<td>Dilution/degradation rate of i (1/s)</td>
</tr>
<tr>
<td>k₁</td>
<td>Half-maximum sugar concentration (M)</td>
<td>K</td>
<td>Half maximum dilution rate (M)</td>
</tr>
<tr>
<td>k₂</td>
<td>Half-maximum sugar concentration (M)</td>
<td>n₁</td>
<td>Hill coefficient for transporter expression</td>
</tr>
<tr>
<td>k₃</td>
<td>Half-maximum regulator concentration (M)</td>
<td>n₂</td>
<td>Hill coefficient for enzyme expression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>v</td>
<td>Intracellular sugar production rate (M/s)</td>
</tr>
</tbody>
</table>

We then defined dimensionless parameters to reduce the total number of parameters. As part of the non-dimensionalization, we use a new time variable (d₀) reflecting the rate of cell division.
### Table S3.2 Dimensionless variables

<table>
<thead>
<tr>
<th>Dimensionless variable</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R'$</td>
<td>$R/R_0$</td>
</tr>
<tr>
<td>$S'$</td>
<td>$S/k_0$</td>
</tr>
<tr>
<td>$S_0'$</td>
<td>$S_0/k_1$</td>
</tr>
<tr>
<td>$T'$</td>
<td>$(d_2/\alpha_3) \cdot T$</td>
</tr>
<tr>
<td>$E'$</td>
<td>$(d_3/\alpha_4) \cdot E$</td>
</tr>
<tr>
<td>$D_1$</td>
<td>$d_1/d_0$</td>
</tr>
<tr>
<td>$D_2$</td>
<td>$d_2/d_0$</td>
</tr>
<tr>
<td>$D_3$</td>
<td>$d_3/d_0$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>$t \cdot d_0$</td>
</tr>
<tr>
<td>$\nu$</td>
<td>$\nu/(k_0d_0)$</td>
</tr>
<tr>
<td>$b_T'$</td>
<td>$d_2/(\alpha_3d_0) \cdot b_T$</td>
</tr>
<tr>
<td>$b_E'$</td>
<td>$d_3/(\alpha_4d_0) \cdot b_E$</td>
</tr>
</tbody>
</table>

Using these dimensionless parameters, equations 1 – 4 were converted into equations 1b-4b as follows:

\[
R' = \frac{S'}{1 + S'} \quad (1b)
\]

\[
\frac{dS'}{d\tau} = \nu + \frac{\alpha_3S_0'}{k_0d_2d_0} + \frac{\alpha_4S'}{k_0d_3d_0} \cdot \frac{1 + S_0'}{k_0d_0} + \frac{S'}{k_0} \cdot \frac{E'}{D_1S'} \quad (2b)
\]

\[
\frac{dT'}{d\tau} = b_T' + D_2 \left( \frac{R'}{R_2} \right)^{\alpha_1} - D_2T' \quad (3b)
\]

\[
\frac{dE'}{d\tau} = b_E' + D_3 \left( \frac{R'}{R_3} \right)^{\alpha_2} - D_3E' \quad (4b)
\]
Finally, using an additional set of dimensionless parameters (Table S3), we derived equations 1c-4c. In this form, the model requires 13 parameter values. By further assuming that basal expression levels are equal ($b_T' = b_E'$) and the intracellular sugar, transporter, and enzyme are diluted at the same rate ($D_1 = D_2 = D_3$), and no endogenous synthesis of sugar ($ν = 0$), the model only requires 9 parameter values.

\[
R' = \frac{S'}{1 + S'} \tag{1c}
\]

\[
\frac{dS'}{dτ} = ν + \frac{S_0'}{1 + S_0'} \alpha T' - \frac{S'}{β + S'} γ E' - D_1 S' \tag{2c}
\]

\[
\frac{dT'}{dτ} = b_T' + D_2 \frac{(R')^{n_1}}{δ^{n_1} + (R')^{n_1}} - D_2 T' \tag{3c}
\]

\[
\frac{dE'}{dτ} = b_E' + D_3 \frac{(R')^{n_2}}{ε^{n_2} + (R')^{n_2}} - D_3 E' \tag{4c}
\]

We selected parameter values within biologically feasible ranges (Table S3.2) to simulate the steady-state behavior of the pathway. The resulting predictions are shown in Figures 3 and 5 as well as Figures S3 and S4. All the simulations of the model were conducted in MATLAB. The ordinary differential equations were integrated using the Euler method. Stable and unstable steady states were calculated by implementing the arc-length continuation method.

In the case where the regulator functions as a repressor, the model is described in equations 1d-4d wherein the parameters and their units are the same as in the previous case:
Non-dimensionalization yields the following set of equations.

\[
R = \frac{R_t}{\frac{k_0}{k_0 + S}}
\]  

(1d)

\[
\frac{dS}{dt} = \nu + \frac{S_0}{k_1 + S_0} T - \frac{S}{k_2 + S} E - \frac{d_3 S}{S}\end{equation}
\]

(2d)

\[
\frac{dT}{dt} = b_T + \frac{\alpha_3 k_4^{n_2}}{k_4 + R^{n_2}} - \frac{d_2 T}{S}
\]  

(3d)

\[
\frac{dE}{dt} = b_E + \frac{\alpha_4 k_4^{n_2}}{k_4 + R^{n_2}} - \frac{d_2 E}{S}\end{equation}
\]

(4d)

Non-dimensionalization yields the following set of equations.

\[
R' = \frac{1}{1 + S'}
\]  

(1e)

\[
\frac{dS'}{d\tau} = \nu + \frac{S_0'}{1 + S_0'} \alpha T' - \frac{S'}{\beta + S'} \gamma E' - D_1 S'
\]  

(2e)

\[
\frac{dT'}{d\tau} = b_T' + D_2 \frac{\delta^{n_1}}{\delta^{n_1} + (R')^{n_1}} - D_2 T'
\]  

(3e)

\[
\frac{dE'}{d\tau} = b_E' + D_3 \frac{\epsilon^{n_2}}{\epsilon^{n_2} + (R')^{n_2}} - D_3 E'
\]  

(4e)

These equations were used in combination with the parameter values in Table S3.3 to generate the bifurcation plots in Figure S3.4.
**Figure S3.1** Measuring pathway expression in single cells by flow cytometry. (A) Overview of generating dot plots. Bacterial cultures in M9 glycerol and the specified concentration of sugar were run directly on the flow cytometer. Cells were gated based on forward scatter and side scatter (red polygon). The gate was selected to collect events from cells but not noisy debris as assessed using DRAQ5 dye (Thermo Scientific). The fluorescence distribution of the sorted events was fit with one or two normal distributions. For two distributions, the lower-fluorescence distribution is called the “uninduced” population and the higher-fluorescence distribution is called the “induced” population. The resulting mean and relative size of each distribution were plotted as dots. The diameter of the dot scales with the fraction of the entire population. (B) Concentration of L-arabinose necessary to induce half of the population (EC$_{50}$) for MG1655 (WT) and MG1655 ΔaraBAD (ΔaraBAD) strains harboring pUA66ParaB following different incubation times. An incubation time of 20 hours was selected as the shortest incubation time to approach a roughly stable distribution. Data points represent the geometric mean and S.E.M. of at least three independent experiments.
**Figure S3.2** Additional response curves for WT MG1655 cells. (A) Response for different promoters. The response was assessed for additional promoters in the D-xylose utilization pathway ($P_{xylF}$) as well as the L-arabinose utilization pathway ($P_{araE}$, $P_{araF}$, $P_{araJ}$). (B) Response to L-arabinose for a single-copy reporter. The $P_{araB}$-$gfp$ construct was inserted into the *E. coli* genome through integration at the $attB$ site. Note that the induced and uninduced sub-populations for $P_{araE}$, $P_{araF}$, and the single integrant of $P_{araB}$ were partially overlapping but generally discernable at intermediate L-arabinose concentrations. (C) Effect of cAMP on the response to L-arabinose. WT cells were grown as described in Figure 3.2, only 2 mM of cAMP was included in the growth medium previously shown to relieve self-catabolite repression by L-arabinose (Katz and Englesberg, 1971). See Figure 3.2 for the experimental growth conditions and an explanation of the dot plots. All dot plots are representative of at least three independent experiments conducted on separate days.
Figure S3.3 Behaviors predicted by the deterministic model for unequal dilution rates of the sugar ($D_1 = 0.74$), transporter ($D_2 = 0.74$), and enzyme ($D_3 = 0.27$). Responses include (A) bistability, (B) excitability, and (C) oscillations. Panels on the left indicate the bifurcation curve under steady-state conditions, where dashed red lines indicate unstable points and dotted blue lines indicate maximal or minimal enzyme levels as part of oscillations. Panels on the right indicate transient simulations at the indicated extracellular sugar concentration ($S_0$). The different colored curves were generated using distinct initial values of intracellular sugar ($S$). Unless stated above, all simulations were conducted using parameter values in Table S3.3. See Figure 3.3 for notations.
Figure S 3.4 Similar behaviors predicted when the regulator functions as a sensory repressor. (A) Major reactions considered by the deterministic model. Extracellular sugar ($S_0$) is imported into the cell by the transporter (T). The intracellular sugar ($S$) can reversibly bind to the regulator or can undergo irreversible degradation by the enzyme (E). The regulator bound to the sugar (R) is unable to repress the expression of the transporter and the enzyme. See page 5 for more details on the model. (B) Bifurcation curves representing the steady-state relationship between enzyme levels and extracellular sugar concentration for a given set of parameter values. Different curves are shown whereby the only altered parameter value is the maximal enzyme activity ($\gamma$). Dashed red lines indicate steady-state enzyme levels that are unstable. See Table S3.3 for values of fixed parameters (with $\nu = 0$).

Figure S3.5 Response curves for $\Delta$xylAB and $\Delta$araBAD strains following excision of the cat resistance cassette. The cassette was removed using the FLP recombinase, resulting in strains MG1655 $\Delta$xylAB -cat and MG1655 $\Delta$araBAD -cat. The curves are identical to those for strains that harbor the cassette, as shown in Figure 3.4. See Figure 3.2 for the experimental growth conditions and an explanation of the dot plots. All dot plots are representative of at least three independent experiments conducted on separate days.
Figure S3.6 Impact of model parameters on the width of the bifurcation region in the absence of catabolism. See Table S3.3 for a brief description of each parameter. Sensitivity analysis was performed with the parameter set in Table S3.3 with $\gamma = 0$. In each simulation, the value of each parameter was either increased (red) or decreased (blue) by two-fold and the resulting change in the width of the bifurcation region is shown. The width of the bifurcation region was measured as the difference between the highest and lowest extracellular sugar concentrations ($S_0$) of the unstable region.
Figure S3.7 The L-arabinose utilization pathway exhibits limited memory of previous conditions, but only in the presence of catabolism. (A) Wild type (WT) and (B) ΔaraBAD (ΔE) cells were grown for 20 hours in M9 glycerol with 0.4 μM or 0.2 μM of L-arabinose, respectively, to induce half of the population (pre). The cells were then sorted into induced and uninduced populations and back diluted into M9 glycerol with 0.15 μM (WT) or 0.05 μM (ΔaraBAD) of L-arabinose, respectively. These concentrations were selected to maximize the difference in response between sorted induced and sorted uninduced cells and to compensate for increased sensitivity to L-arabinose at later incubation times (Figure S1). The back-diluted cultures then were grown for the indicated time prior to flow cytometry analysis. The fraction of the population that is induced at each time point is shown on the right.
**Supplementary Tables**

**Table S3.3** Overview of utilization pathways with inducibly expressed transporters and catabolic enzymes in *E. coli*. Some of the sugars can be transported through other transport systems, although we only list systems induced by the imported sugar. The displayed information is from Ecocyc ([www.ecocyc.org](http://www.ecocyc.org)). aProteins separated by slashes participate in a single complex (e.g. AraF/G/H form the high-affinity transport system for L-arabinose) or a linear cascade (e.g. AraB/A/D shunt L-arabinose into the pentose phosphate pathway through three linear reactions). bSome sugars are phosphorylated upon entering the cell through phosphotransferase systems, although the phosphorylation event occurs upstream of catabolic events of the enzyme(s). cPathways can catabolize the sugar (Cat) or both catabolize and synthesize the sugar (Cat/Syn).

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Regulators</th>
<th>Transporters</th>
<th>Enzymes</th>
<th>Inducers</th>
<th>Pathway type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GntT, GntU, GntP, GntW</td>
<td>GntK, IdnK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-gluconate</td>
<td>GntR</td>
<td></td>
<td></td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-sorbitol</td>
<td>GutM, GutR</td>
<td>SrlA/B/E</td>
<td>SrlD</td>
<td>Imported sugar</td>
<td>Cat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-xylose</td>
<td>XylR</td>
<td>XylE, XylF/G/H</td>
<td>XylA/B</td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-fructose</td>
<td>Cra</td>
<td>FruA/B</td>
<td>FruK</td>
<td></td>
<td>Cat/Syn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-arabinose</td>
<td>AraC</td>
<td>AraE, AraF/G/H</td>
<td>AraA/B/ D</td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td>L-fucose</td>
<td>FucR</td>
<td>FucP</td>
<td>FucI/K/A/O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-iodonate</td>
<td>IdnR</td>
<td>IdnT</td>
<td>IdnD/O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-lactate</td>
<td>LldR</td>
<td>LldP</td>
<td>LldD</td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td>L-fructose</td>
<td>XylR</td>
<td>XylE, XylF/G/H</td>
<td>XylA/B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GntT, GntU, GntP, GntW</td>
<td>GntK, IdnK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-gluconate</td>
<td>GntR</td>
<td></td>
<td></td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-sorbitol</td>
<td>GutM, GutR</td>
<td>SrlA/B/E</td>
<td>SrlD</td>
<td>Imported sugar</td>
<td>Cat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-xylose</td>
<td>XylR</td>
<td>XylE, XylF/G/H</td>
<td>XylA/B</td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-fructose</td>
<td>Cra</td>
<td>FruA/B</td>
<td>FruK</td>
<td></td>
<td>Cat/Syn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-arabinose</td>
<td>AraC</td>
<td>AraE, AraF/G/H</td>
<td>AraA/B/ D</td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td>L-fucose</td>
<td>FucR</td>
<td>FucP</td>
<td>FucI/K/A/O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-iodonate</td>
<td>IdnR</td>
<td>IdnT</td>
<td>IdnD/O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-lactate</td>
<td>LldR</td>
<td>LldP</td>
<td>LldD</td>
<td>Sugar</td>
<td>Cat</td>
</tr>
<tr>
<td>L-fructose</td>
<td>XylR</td>
<td>XylE, XylF/G/H</td>
<td>XylA/B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

100
Table S3.4 Doubling times in the presence or absence of sugar. Doubling times were measured for the indicated strain growing in exponential phase (ABS\textsubscript{600} of 0.01 – 0.3) in M9 glycerol without (-S) or with (+S) 1 mM of the indicated sugar. Values represent the average ± S.E.M. for three independent experiments. *Not determined. At higher turbidities (ABS\textsubscript{600} > 0.01), the ΔnagB strain exhibited greatly diminished growth in the presence of 1 mM of N-acetylglucosamine. However, at the lower turbidities in which the flow cytometry analysis was conducted (Figure 3.4), the ΔnagB strain and the WT strain appeared to grow at comparable rates.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Reporter</th>
<th>Sugar</th>
<th>Doubling time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-S</td>
<td>+S</td>
</tr>
<tr>
<td>WT</td>
<td>pUA66ParaB</td>
<td>L-arabinose</td>
<td>129 ± 8</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td></td>
<td></td>
<td>80 ± 9</td>
</tr>
<tr>
<td>WT</td>
<td>pUA66PxyIA</td>
<td>D-xylose</td>
<td>88 ± 9</td>
</tr>
<tr>
<td>ΔxylAB</td>
<td></td>
<td></td>
<td>82 ± 2</td>
</tr>
<tr>
<td>WT</td>
<td>pUA66rhaB</td>
<td>L-rhamnose</td>
<td>107 ± 11</td>
</tr>
<tr>
<td>ΔrhaB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>pUA66galE</td>
<td>D-galactose</td>
<td>105 ± 4</td>
</tr>
<tr>
<td>ΔgalKM</td>
<td></td>
<td></td>
<td>85 ± 5</td>
</tr>
<tr>
<td>WT</td>
<td>pUA66nagB</td>
<td>N-acetylglucosamine</td>
<td>88 ± 6</td>
</tr>
<tr>
<td>ΔnagB</td>
<td></td>
<td></td>
<td>98 ± 1</td>
</tr>
<tr>
<td>WT</td>
<td>pUA66nanA</td>
<td>N-acetylneuraminic</td>
<td>116 ± 11</td>
</tr>
<tr>
<td>ΔnanA</td>
<td></td>
<td>acid</td>
<td>90 ± 6</td>
</tr>
<tr>
<td>WT</td>
<td>pUA66gntK</td>
<td>D-gluconate</td>
<td>92 ± 6</td>
</tr>
<tr>
<td>WT</td>
<td>pUA66lacZ</td>
<td>D-lactose</td>
<td>112 ± 13</td>
</tr>
</tbody>
</table>
Table S3.5 Parameters for the model of sugar utilization with a sensory activator or repressor. The potential biological ranges for each parameter were calculated based on order-of-magnitude estimates for each ratio. For instance, the estimate of $b_T'$ was estimated based on the dynamic ranges reported for promoters of transporters. Extended ranges were purposefully used for some dimensionless parameters because of widely ranging values reported in the literature (e.g. relative binding affinity of catabolic enzymes) or a lack of any values (e.g. rates of endogenous production). All model predictions used the sensory activator except for those in Figure S3.4, which used the sensory repressor.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Dimensionless parameter</th>
<th>Value</th>
<th>Potential biological range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\alpha_1 \alpha_3}{k_0 d_2 d_0}$</td>
<td>$\alpha$</td>
<td>See main text</td>
<td>$R = \text{activator}$</td>
</tr>
<tr>
<td>$\frac{k_2}{k_0}$</td>
<td>$\beta$</td>
<td>$0.0005$</td>
<td>$10$</td>
</tr>
<tr>
<td>$\frac{\alpha_2 \alpha_4}{k_0 d_2 d_0}$</td>
<td>$\gamma$</td>
<td>See main text</td>
<td>$R = \text{repressor}$</td>
</tr>
<tr>
<td>$\frac{k_3}{R_t}$</td>
<td>$\delta$</td>
<td>$0.0004$</td>
<td>$0.1$</td>
</tr>
<tr>
<td>$\frac{k_4}{R_t}$</td>
<td>$\varepsilon$</td>
<td>$0.0005$</td>
<td>$0.1$</td>
</tr>
<tr>
<td>$\frac{b_T D_2}{\alpha_3}$</td>
<td>$b_T'$</td>
<td>$0.00033$</td>
<td>$0.00033$</td>
</tr>
</tbody>
</table>
Table S3.5 Continued

<table>
<thead>
<tr>
<th>$\frac{D_3}{d_0}$</th>
<th>$b_{\theta}^\prime$</th>
<th>$0.00033$</th>
<th>$0.00033$</th>
<th>$0.0001 – 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>$n_1$</td>
<td>$2$</td>
<td>$2$</td>
<td>$0.5 – 5$</td>
</tr>
<tr>
<td>$n_2$</td>
<td>$n_2$</td>
<td>$1.5$</td>
<td>$2$</td>
<td>$0.5 – 5$</td>
</tr>
<tr>
<td>$\frac{d_1}{d_0}$</td>
<td>$D_1$</td>
<td>$0.74$</td>
<td>$0.74$</td>
<td>$0.001 – 1$</td>
</tr>
<tr>
<td>$\frac{d_2}{d_0}$</td>
<td>$D_2$</td>
<td>$0.74$</td>
<td>$0.74$</td>
<td>$0.001 - 1$</td>
</tr>
<tr>
<td>$\frac{d_3}{d_0}$</td>
<td>$D_3$</td>
<td>$0.74$</td>
<td>$0.74$</td>
<td>$0.001 - 1$</td>
</tr>
<tr>
<td>$\nu'$</td>
<td>$\nu$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0 - 1$</td>
</tr>
</tbody>
</table>
Table S3.6 Strains used in this work.

<table>
<thead>
<tr>
<th>Name</th>
<th>Genotype</th>
<th>Stock #</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td><em>Escherichia coli</em> MG1655</td>
<td>CB168</td>
</tr>
<tr>
<td>NM500</td>
<td>MG1655 mini-λ-tetA</td>
<td>CB26</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>MG1655 ΔaraBAD::cat</td>
<td>CB225</td>
</tr>
<tr>
<td>ΔaraBAD -cat</td>
<td>MG1655 ΔaraBAD</td>
<td>CB387</td>
</tr>
<tr>
<td>ΔxylAB</td>
<td>MG1655 ΔxylAB::cat</td>
<td>CB226</td>
</tr>
<tr>
<td>ΔxylAB -cat</td>
<td>MG1655 ΔxylAB</td>
<td>CB388</td>
</tr>
<tr>
<td>ΔgalKM</td>
<td>MG1655 ΔgalKM::cat</td>
<td>CB259</td>
</tr>
<tr>
<td>ΔrhaBADM</td>
<td>MG1655 ΔrhaBADM::cat</td>
<td>CB260</td>
</tr>
<tr>
<td>ΔnanA</td>
<td>MG1655 ΔnanA</td>
<td>CB335</td>
</tr>
<tr>
<td>ΔnagB</td>
<td>MG1655 ΔnagB</td>
<td>CB336</td>
</tr>
<tr>
<td>attB::P&lt;sub&gt;ara&lt;/sub&gt;-gfp</td>
<td>MG1655 attB::P&lt;sub&gt;ara&lt;/sub&gt;-gfp</td>
<td>CB389</td>
</tr>
</tbody>
</table>
Table S3.7 Plasmids used in this work.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Source</th>
<th>Stock #</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUA66</td>
<td>Low-copy plasmid (pSC101 ori) with MCS upstream of gfp</td>
<td>OpenBiosystems</td>
<td>pCB198</td>
</tr>
<tr>
<td>pUA66ParaB</td>
<td>pUA66 with the araB promoter</td>
<td>OpenBiosystems</td>
<td>pCB208</td>
</tr>
<tr>
<td>pUA66ParaE</td>
<td>pUA66 with the araE promoter</td>
<td>This study</td>
<td>pCB286</td>
</tr>
<tr>
<td>pUA66ParaF</td>
<td>pUA66 with the araF promoter</td>
<td>This study</td>
<td>pCB287</td>
</tr>
<tr>
<td>pUA66ParaJ</td>
<td>pUA66 with the araJ promoter</td>
<td>OpenBiosystems</td>
<td>pCB288</td>
</tr>
<tr>
<td>pUA66PxyI</td>
<td>pUA66 with the xylA promoter</td>
<td>This study</td>
<td>pCB289</td>
</tr>
<tr>
<td>pUA66PxyIF</td>
<td>pUA66 with the xylF promoter</td>
<td>This study</td>
<td>pCB290</td>
</tr>
<tr>
<td>pUA66PntK</td>
<td>pUA66 with the gntK promoter</td>
<td>OpenBiosystems</td>
<td>pCB291</td>
</tr>
<tr>
<td>pUA66PnagB</td>
<td>pUA66 with the nagB promoter</td>
<td>OpenBiosystems</td>
<td>pCB292</td>
</tr>
<tr>
<td>pUA66PnagB</td>
<td>pUA66 with the rhaB promoter</td>
<td>OpenBiosystems</td>
<td>pCB293</td>
</tr>
<tr>
<td>pUA66PnagB</td>
<td>pUA66 with the nanA promoter</td>
<td>OpenBiosystems</td>
<td>pCB337</td>
</tr>
<tr>
<td>pUA66PnagB</td>
<td>pUA66 with the galE promoter</td>
<td>OpenBiosystems</td>
<td>pCB338</td>
</tr>
<tr>
<td>pUA66PnagB</td>
<td>pUA66 with the lacZ promoter</td>
<td>OpenBiosystems</td>
<td>pCB44</td>
</tr>
<tr>
<td>pCP20</td>
<td>Heat-inducible expression of the FLP recombinase</td>
<td>CGSC #pCP20</td>
<td>pCB390</td>
</tr>
<tr>
<td>pCAH63</td>
<td>CRIM plasmid</td>
<td>CGSC #7866</td>
<td>pCB390</td>
</tr>
<tr>
<td>pINT-ts</td>
<td>Helper integration plasmid</td>
<td>CGSC #7855</td>
<td>pCB391</td>
</tr>
<tr>
<td>pCAH63-ParaB</td>
<td>CRIM plasmid with $P_{ara}$-gfp in place of the $uidA$ stuffer</td>
<td>This study</td>
<td>pCB392</td>
</tr>
</tbody>
</table>
Table S3.8 Oligonucleotides used in this work. Sequences in bold denote overlap regions for Gibson cloning assembly. Sequences in green denote restriction sites used to clone the amplified region into pUA66. Underlined sequences denote homology regions for recombineering.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ParaE.fwd</td>
<td>TATCACGAGGCCCTTTTCGTTCCTCACCTGATGGGGCCGATAGTG</td>
</tr>
<tr>
<td>ParaE.rev</td>
<td>TGTATAATCTCCTCTCTTTAATCCTAGATAACGCAAGCAGCTTCAGGA</td>
</tr>
<tr>
<td>ParaF.fwd</td>
<td>TATCACGAGGCCCTTTTCGTTCCTCACGAAGCATTCACAGCGGTTG</td>
</tr>
<tr>
<td>ParaF.rev</td>
<td>TGTATAATCTCCTCTTTAATCCTAGATAACGAGCTTCAGGGTCTCC</td>
</tr>
<tr>
<td>PxyIA.fwd</td>
<td>TATCACGAGGCCCTTTTCGTTCCTCACGGTGTAAGGGCCCTTCGTTCAGGTAT</td>
</tr>
<tr>
<td>PxyIA.rev</td>
<td>TGTATAATCTCCTCTTTAATCCTAGATACATACGCTACCCAAACACCA</td>
</tr>
<tr>
<td>PxyIF.fwd</td>
<td>AATACTCGAGTTACTGTCGAAATAGGCTTGGCATATTTG</td>
</tr>
<tr>
<td>PxyIF.rev</td>
<td>AATAAGATCCTGCGTCAAGGCCCCTTCGTTCAGGTAT</td>
</tr>
<tr>
<td>pUA66.fwd1</td>
<td>TCTAGATTTAAGAAGGAGATATACATATGAGTAAGGAG</td>
</tr>
<tr>
<td>pUA66.rev1</td>
<td>ATGCTTTATGGTGTAATGCTAAAGCAA</td>
</tr>
<tr>
<td>pUA66.fwd2</td>
<td>TTAGTTTAGTAATACACATAGCATAGCATATGAGTTG</td>
</tr>
<tr>
<td>pUA66.rev2</td>
<td>GTGAAGACGAAAGGCCCCTCG</td>
</tr>
<tr>
<td>del-araBAD.fwd</td>
<td>TCCATACCCGTCTTTTTTTGAGATGGAAACGCAATGGTCCATATGATA</td>
</tr>
<tr>
<td>del-araBAD.rev</td>
<td>GTTGTGCTGATATCCGCTGACTGCGCTAAACTGCAAGGCGATAGGTAGCT</td>
</tr>
<tr>
<td>del-xylAB.fwd</td>
<td>ACCCGGCCCATTACCTGTATTGGAGTTCAATGTGCTCAATGAAATA</td>
</tr>
<tr>
<td>del-xylAB.rev</td>
<td>GGAAGAACCAGTCAATGAGCGATTCTTGAGTTCTGAGCTCGATAGGTAGCT</td>
</tr>
<tr>
<td>del-nanA.fwd</td>
<td>AAGGTATACGTGGTTACTGAGCAAGCATACTTTGAGGATTGTAAGGTC</td>
</tr>
<tr>
<td>del-nanA.rev</td>
<td>TGAAGACTGCTT</td>
</tr>
</tbody>
</table>
**Table S3.8 Continued**

<table>
<thead>
<tr>
<th>Segment</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>del-nagB.fwd</td>
<td>CAGGTGGATTGACGCTGTCTGCAGCAATCCAGGTTACTGCTTGAGCC</td>
</tr>
<tr>
<td>del-nagB.rev</td>
<td>GGAACATTTCGGTGCCAGGGTGTAATGACGTCATATGAAATCCCTCTCTTCTGGAGCTGCTT</td>
</tr>
<tr>
<td>del-galKM.fwd</td>
<td>ATCCATTTTCGGAATCCGGGAGTGAAGAATACGTTCCATATGAAATAATCCCTCTCTTCTGGAGCTGCTT</td>
</tr>
<tr>
<td>del-galKM.rev</td>
<td>TGGGACGAAGCAGGAAATACGGGAGAAGGAAGTAAGTCGACGTAATGACGTCATATGAAATA</td>
</tr>
<tr>
<td>del-rhaBADM.fwd</td>
<td>CGCTTTTTAGACTGTCGTATGAAATACGTTCCATATGAAATAATCCCTCTCTTCTGGAGCTGCTT</td>
</tr>
<tr>
<td>del-rhaBADM.rev</td>
<td>ATGGTGCGCCAGGAGAGTTGTTGATTTGTATAGTTCATCCATGCC</td>
</tr>
<tr>
<td>attB-ParaB.fwd</td>
<td>ATGGTGCGCCAGGAGAGTTGTTGATTTGTATAGTTCATCCATGCC</td>
</tr>
<tr>
<td>attB-gfp.rev</td>
<td>ATGGTGCGCCAGGAGAGTTGTTGATTTGTATAGTTCATCCATGCC</td>
</tr>
<tr>
<td>attB_plasmid.fwd</td>
<td>ATCAACAACTCTCTGTGGCGCACCAT</td>
</tr>
<tr>
<td>attB_plasmid.rev</td>
<td>TCTTGAAGACACTAGTTCGGGACAG</td>
</tr>
</tbody>
</table>

**Supplementary References**


Chapter 4

Trade-offs in engineering sugar utilization pathways for titratable control

*This work has been published in *ACS Synthetic Biology*. Contributions to this work: Kelsey Boykin, Konstantinos Biliouris, Yiannis Kaznessis
Abstract

Titratable systems are common tools in metabolic engineering to tune the levels of enzymes and cellular components as part of pathway optimization. For non-model microorganisms with limited genetic tools, inducible sugar utilization pathways offer built-in titratable systems. However, these pathways can exhibit undesirable single-cell behaviors that hamper the uniform and tunable control of gene expression. Here, we applied mathematical modeling and single-cell measurements of the L-arabinose utilization in *Escherichia coli* to systematically explore how sugar utilization pathways can be altered to achieve desirable inducible properties. We found that different pathway alterations, such as the removal of catabolism, constitutive expression of high-affinity or low-affinity transporters, or further deletion of the other transporters, came with trade-offs specific to each alteration. For instance, sugar catabolism improved the uniformity and linearity of the response at the cost of requiring higher sugar concentrations to induce the pathway. Within these alterations, we also found that a uniform and linear response could be achieved with a single-alteration: constitutively expressing the high-affinity transporter. Equivalent modifications to the D-xylose utilization pathway yielded similar responses, demonstrating the applicability of our observations. Overall, our findings indicate that there is no ideal set of typical alterations when co-opting natural utilization pathways for titratable control and suggest design rules for manipulating these pathways to advance basic genetic studies and the metabolic engineering of microorganisms for optimized chemical production.
4.1 Introduction
Metabolic engineering requires the optimization of native and heterologous pathways in order to maximize the yield of the desired chemical product, prevent the build-up of toxic intermediates, and maintain cell viability. A common strategy is finely tuning the expression of each cellular component using titratable systems.\textsuperscript{1-3} These systems rely on exogenous inducer molecules to tune the expression levels of regulated genes; by varying the concentration of the inducer, the expression levels of the gene can be manipulated without additional genetic modifications to the host organism. A number of titratable systems are available that were derived from or inspired by natural processes—including nutrient utilization pathways,\textsuperscript{2-5} riboswitches,\textsuperscript{6-9} antibiotic resistance cassettes,\textsuperscript{10-12} quorum sensing,\textsuperscript{13} and light sensing.\textsuperscript{14}

Metabolic engineering has been more frequently adopting non-model organisms that already possess desired pathways or have adapted to environments that parallel production conditions.\textsuperscript{15} Achieving titratable control in these organisms is more challenging because of limited genetic tools as well as growth conditions that perturb the stability and function of the inducers or regulators. However, these organisms often encode multiple sugar utilization pathways that can serve as ready-made titratable systems.\textsuperscript{16-25} Sugar utilization pathways are generally comprised of transporters that import the sugar into the cytoplasm, catabolic enzymes that shunt the sugar into central metabolism, and regulators that activate the expression of the transporters and the enzymes when bound to the sugar (Figure 4.1A). Furthermore, pathways often respond to their individual sugar, offering orthogonal systems for the coordinated optimization of gene expression.

The drawback of co-opting natural utilization pathways is their potential to exhibit undesirable behaviors. While an ideal titratable response should be uniform and linear (Figure 4.1B), sugar utilization pathways have been shown to exhibit bimodal, sharp responses that prevent fine-tuning of expression levels in all cells of the population.\textsuperscript{26-30} These responses can be attributed to positive feedback within the pathway, which emerges from the sugar inducing the transporters that import more sugar into the cell.\textsuperscript{27,29,31,32} The pathways also possess negative feedback from induction of the catabolic enzymes that break
down the sugar.\textsuperscript{33} While negative feedback would be expected to counterbalance positive feedback\textsuperscript{33,34}, sugar catabolism also degrades the inducing sugar. With all of these contributing factors, a fundamental question is how to modify a given pathway to achieve a desirable titratable response.

![Figure 4.1](image)

**Figure 4.1.** Co-opting sugar utilization pathways for titratable control of gene expression. (A) Components of inducible sugar utilization pathways. High affinity (T\textsubscript{H}) and low-affinity (T\textsubscript{L}) transporters import the sugar into the cell, while catabolic enzymes (E) shunt the sugar into central metabolism. Transcription regulators (R) up-regulate the expression of the transporters and the enzymes, but only in the presence of the sugar. (B) Desirable properties for a titratable response. These properties include a uniform response at all inducer concentrations, a large dynamic range (high \(\delta\)), low inducer concentrations to induce the pathway (low EC\textsubscript{50}), and strong linearity (low \(\eta\)).

A powerful model has been the L-arabinose utilization pathway in *Escherichia coli*. This pathway encodes one high-affinity transporter (encoded within the *araFGH* operon) and one low affinity transporter (encoded by the *araE* gene), is well characterized, and has been regularly used for inducible control. This pathway has also been shown to exhibit a bimodal response, complicating its use as a titratable system.\textsuperscript{26} Accordingly, previous efforts sought to eliminate bimodality through the overexpression of one transporter along with multiple alterations to the pathway. Khlebnikov and coworkers demonstrated that heterologously
expressing the low-affinity transporter AraE could remove bimodality, which required disrupting both endogenous transporters.\textsuperscript{35,36} Separately, Morgan-Kiss and coworkers demonstrated that heterologously expressing a substrate-relaxed mutant of the lactose transporter could eliminate bimodality in a strain lacking the L-arabinose transporters and catabolic enzymes.\textsuperscript{37} As a result, the consensus is that overexpressing a transporter and stripping away almost all native components of the pathway should be performed to achieve titratable control. While these studies laid the groundwork for generating utilization pathways with titratable responses, they called for extensive genetic manipulations (i.e. plasmid-based expression of a transporter, disruption of multiple genetic loci) that may be difficult to achieve in non-model organisms and may not reflect ideal configuration for titratable control. In addition, a systematic analysis of all pathway alterations remains to be performed, which could reveal other configurations that require fewer modifications or generate a more desirable titratable response.

Here, we systematically evaluated how natural utilization pathways can be altered to achieve titratable control. Mathematical modeling and experimental probing of the L-arabinose utilization pathway revealed distinct trade-offs when altering the pathway: (i) constitutively expressing the low affinity-transporter yielded the most linear response but required additional modifications to eliminate bimodality, (ii) constitutively expressing one transporter and deleting the other transporter had both positive and negative effects on the response properties specific to the selected transporter, and (iii) sugar catabolism linearized the response and reduced the extent of bimodality at the cost of elevated sugar concentrations to induce the pathway. Within these trade-offs, one of the most desirable responses came from a single alteration: constitutively expressing the high-affinity transporter. Furthermore, maintaining sugar catabolism improved the linearity of the response at higher cell densities, perceivably due to negative feedback through depletion of the exogenous sugar. Similar trends were observed for the D-xylose utilization pathway, suggesting broad applicability of our findings. These insights demonstrate that there is no perfect set of typical alterations when co-opting sugar utilization pathways, wherein the optimal configuration will depend on the specific demands placed on the titratable system.
4.2 Results and discussion
We considered four metrics to assess the desirability of the response (Figure 4.1B): the extent of bimodality (the number of peaks in the fluorescence distribution), the dynamic range of the response ($\delta$, measured as the ratio of expression levels at full and no induction), the inducer concentration to achieve half-maximal induction ($EC_{50}$), and the steepness of the response curve ($\eta$, measured as a Hill coefficient). An ideal titratable system would exhibit uniform expression across the cell population at all inducer concentrations, a large dynamic range, a low inducer concentration to achieve half-maximal induction, and a low Hill coefficient reflecting the steepness of the response curve. We used these metrics as the basis to evaluate the predicted or measured behaviors of sugar utilization.

4.2a A simple mathematical model of sugar utilization
We built a simple mathematical model of sugar utilization to qualitatively explore how different alterations to the pathway influence the response properties (see Supporting Information). This model was comprised of two transporters that import sugar into the cell, one enzyme that breaks down the sugar, and one regulator that binds the sugar and subsequently activates the expression of the transporters and the enzyme. This general configuration captures the L-arabinose utilization pathway and many other pathways found in microorganisms.$^{16-25}$ Parameter values were selected for one high-affinity/low-capacity transporter and one low-affinity/high-capacity transporter at biologically relevant expression strengths and activities to yield a bistable response (Figure 4.2, see Supporting Information). This configuration parallels the behavior of the L-arabinose utilization pathway and creates an undesired behavior for titratable control.

4.2b Experimentally probing L-arabinose utilization to explore model predictions
We concurrently altered and assessed the L-arabinose utilization pathway in *E. coli* K-12. This pathway was attractive because of its extensive characterization,$^{38}$ its widespread use as a titratable system,$^{2}$ and its bimodal response that prevents titratable control.$^{26,35,37}$ The pathway comprises three enzymes (encoded by *araBAD*) that shunt L-arabinose into the
pentose phosphate pathway, a low-affinity transporter (encoded by *araE*), a high-affinity transporter (encoded by *araFGH*), and a dual transcription regulator (encoded by *araC*) that upregulates the expression of all other pathway components when bound to L-arabinose (Figure S4.1A). A third L-arabinose-responsive gene (*araJ*) encodes a putative transporter, although this gene appears to contribute negligibly to the pathway response\(^{39,40}\) and thus was not considered in our analyses.

To measure the response to L-arabinose, we equipped the wild type (WT) strain of *E. coli* MG1655 with a low-copy plasmid encoding the *araB* promoter upstream of the green fluorescent protein (GFP) gene.\(^{41}\) The associated strain was then grown for 6 hours in a defined medium supplemented with different concentrations of L-arabinose to mid-log phase.
Figure 4.2. A simple mathematical model predicts trade-offs when altering the pathway structure. (A) The model assumes a base pathway comprising a high-affinity/low-capacity transporter ($T_H$) and a low-affinity/high-capacity transporter ($T_L$) that import extracellular sugar ($S_0$) into the cell, a catabolic enzyme (E) that degrades the sugar, and a constitutively expressed regulator that upregulates the expression of the transporters and the enzymes when bound to the sugar. The steady-state expression levels of the enzyme are reported as a function of extracellular sugar concentration. Note that all variables were non-dimensionalized as part of the model derivation. Dashed lines indicate bifurcation regions. To alter the pathway, $T_H$ was constitutively expressed (C,D) and the activity of $T_L$ was further eliminated (E,F), $T_L$ was constitutively expressed (G,H) and the activity of $T_H$ was further eliminated (I,J), and the activity of the catabolic enzymes was eliminated (B,D,F,H,J). Three strengths of constitutive expression were selected for $T_H$ and $T_L$ (low, light blue; medium, blue; high, dark blue). See Supporting Information for more details.
(ABS<sub>600</sub> ~0.4) and subjected to flow cytometry analysis (Figure S4.2). The incubation time was selected such that longer incubation times would yield similar fluorescence distributions (Figure S4.3). Figure 4.3 displays representative response curves while Table 4.1 reports the calculated response metrics. The response curves are depicted as dot plots to capture the extent of bimodality within each fluorescence distribution (the relative size of each dot) and the fluorescence of each subpopulation (vertical location of each dot). Following this approach, the WT strain yielded a bimodal response across a wide range of L-arabinose concentrations (Figure 4.3A, Table 4.1) similar to previous observations.<sup>35,37</sup>

4.2c Constitutively expressing the high-affinity transporter more readily generates a uniform response

We considered different alterations to sugar utilization pathways that were previously applied: constitutively expressing one of the transporters, deleting the other transporter, and deleting the catabolic genes.<sup>35–37</sup> From a network perspective, each alteration affects feedback within the pathway: constitutively expressing or deleting the transporter genes disrupts positive feedback, whereas deleting the catabolic genes disrupts negative feedback. While differential growth on the sugar could also impart feedback,<sup>42</sup> we found that the addition of L-arabinose did not have a major impact on growth rates for any of the pathway alterations (Table S4.1).

We first explored the impact of constitutively expressing either transporter to partially disrupt positive feedback in the pathway. Within the model, we replaced the regulator-dependent expression term for either transporter with different constant values. We found that all evaluated expression strengths of the high-affinity transporter eliminated the bifurcation region (Figure 4.2C). In contrast, only the highest expression strength of the low-affinity transporter eliminated the bifurcation region (Figure 4.2G). In both cases, increasing the promoter strength reduced the EC<sub>50</sub> value by allowing greater sugar import at lower extracellular concentrations. Because positive feedback is required for bistability, the model predictions suggest that the high-affinity transporter rather than the low-affinity transporter primarily drives bistability, at least under the selected parameter set.
To experimentally probe model predictions, we replaced the native \textit{araF} or \textit{araE} promoter with a medium-strength, synthetic promoter (P\textsubscript{con-araFGH} or P\textsubscript{con-araE}) and measured the response to L-arabinose (Figure 4.3C,G). Paralleling model predictions, constitutively expressing the high-affinity transporter eliminated bimodality, whereas constitutively expressing the \textit{araE} low-affinity transporter reduced but did not eliminate bimodality. These results clearly demonstrate that (i) constitutively expressing the high-affinity transporter \textit{araFGH} more readily generated a uniform response and (ii) only a single modification was required to fully convert a bimodal response into a uniform response.

\textbf{4.2d Trade-offs when deleting one transporter}

We next explored the impact of deleting one of the transporters when the other transporter was constitutively expressed. Within the model, the second transporter was removed by setting its maximal activity ($\alpha_H$, $\alpha_L$) equal to zero. The resulting impact on the predicted response was mixed and specific to the transporter (Figure 4.2E,I). Removing the high-affinity transporter from the pathway with a constitutively expressed low-affinity transporter eliminated the bifurcation region but also shifted the curves to higher extracellular sugar concentrations (Figure 4.2I). This effect may be attributed to the higher-affinity transporter allowing sugar import at lower sugar concentrations while driving positive feedback within the pathway. In support of model predictions, deletion of the \textit{araFGH} operon (Δ\textit{araFGH}) from the L-arabinose pathway with constitutively expressed \textit{araE} resulted in a uniform response with a modest but statistically significant increase in the EC\textsubscript{50} value (p-value = 0.019) (Figure 4.3I, Table 4.1).

The model predicted that removing the low-affinity transporter from the pathway constitutively expressing the high-affinity transporter had a negligible effect (compare Figures 4.2C and 4.2E). While this may suggest that the low-affinity transporter is poorly induced for our parameter set, the low-affinity transporter alone could mediate full induction of the pathway at high extracellular sugar concentrations (Figure S4.4). Deletion of \textit{araE} (Δ\textit{araE}) from the L-arabinose utilization pathway constitutively expressing \textit{araFGH} also maintained the uniform response (Figure 4.3E). However, there was a significant decrease in
the EC$_{50}$ value (p-value = 0.008) and a significant increase in the Hill coefficient (p-value = 0.003). These changes would suggest that the low-affinity transporter contributes more than that predicted by the model by extending the concentration range that titrates the response. These changes also suggest that deletion of the low-affinity transporter poses a distinct trade-off: a lower EC$_{50}$ value at the cost of a sharper response.

4.2e Trade-offs when eliminating sugar catabolism

We next explored the impact of removing sugar catabolism. Previous work concluded that catabolism posed a barrier to a titratable response by breaking down the inducing molecule.$^{37}$ Separately, sugar catabolism was predicted to entirely eliminate bistability,$^{33}$ suggesting conflicting contributions. To examine the impact of removing catabolism in the model, we set the activity of the enzyme equal to zero ($\alpha_E = 0$) and evaluated the impact on the response metrics for all pathway configurations (Figure 4.2B,D,F,H,J). The model predicted both positive and negative effects of removing catabolism. As a positive effect, removing sugar catabolism lowered the apparent EC$_{50}$ value by preventing the intracellular breakdown of the sugar. As two negative effects, removing sugar catabolism enhanced the propensity for bistability and sharpened the response. Both negative effects can be attributed to sugar catabolism conferring negative feedback on the pathway similar to negative feedback in transcriptional circuits.$^{34,43,44}$

To test these predictions, we deleted the araBAD operon (ΔaraBAD) from all strains and measured the single-cell response curves (Figure 4.3B,D,F,H,J, Table 4.1). For all pathways (Figure 4.3), the response metrics aligned with model predictions. In particular, removing L-arabinose catabolism significantly lowered the EC$_{50}$ value (p-values = 1x10$^{-6}$ – 0.008), with over 100-fold differences in some cases.
Figure 4.3. Probing alterations to the L-arabinose utilization pathway in *E. coli*. The wild type pathway (A) was subjected to different alterations: *araFGH* was constitutively expressed (P\textsubscript{con-araFGH}) (C,D) and *araE* was further deleted (Δ*araE*) (E,F), *araE* was constitutively expressed (P\textsubscript{con-araE}) (G,H) and *araFGH* was further deleted (Δ*araFGH*) (I,J), and *araBAD* was deleted (Δ*araBAD*) (B,D,F,H,J). Each designated strain was back-diluted into M9 minimal medium supplemented with the indicated concentration of L-arabinose and grown for 6 hours to ABS\textsubscript{600} ~0.4 prior to flow cytometry analysis. For unimodal distributions, the resulting mean fluorescence is plotted. For bimodal distributions, two dots are plotted to represent the mean fluorescence and the relative number of cells in the induced (black) and uninduced (white) subpopulations (see Figure S4.2 for more details on the flow cytometry analysis). The diameter of each dot is directly proportional to the fraction of cells in that subpopulation. Gray boxes indicate the limit of detection due to autofluorescence. Each dot plot is representative of at least three independent experiments conducted on separate days. See Table 4.1 for the response metrics that account for the replicate experiments.
Furthermore, removing L-arabinose catabolism sharpened the response (p-values = 3x10^{-5} – 0.005) and increased the propensity for bimodality. Finally, removing L-arabinose catabolism did not have a consistent impact on the dynamic range, where a statistically significant increase was observed for only two strains (P_{con-araFGH}, p-value = 0.027; P_{con-araFGH ΔaraE}, p-value = 0.010) (Table 4.1). Thus, we conclude that removing L-arabinose catabolism poses a major tradeoff between the amount of sugar necessary to induce the pathway and the sharpness of the response.

**4.2f Breakdown of the sugar can help linearize the response**

Catabolism can deplete the reserve of sugar in the medium, which has been viewed as a barrier to the generation of titratable systems. However, such depletion would form negative feedback by depleting the available sugar to induce the pathway, potentially benefiting the response. To explore these potential effects, we first modified the simple model to allow for the depletion of the extracellular sugar through catabolism (see Supporting Information). The resulting model was then employed to predict the response in the presence of catabolism. As part of the model, we could specify the relative volume of the cells to the medium (ν), which dictates in part the rate of depletion. We specifically focused on the pathway with the constitutively expressed low-affinity transporter (T_L = 0.02) and the deleted high-affinity transporter (α_H = 0) that yielded a graded response in the presence or absence of sugar catabolism. The model predicted that depletion of the extracellular sugar flattened the curve and elevated the EC_{50} value (Figure 4.4A). High depletion (ν = 0.01) was detrimental based on the sharp response close to saturating sugar concentrations, although intermediate depletion (ν < 0.01) improved the overall linearity of the response.

To test these predictions, we cultured the P_{con-araFGH ΔaraE} and P_{con-araFGH ΔaraE ΔaraBAD} strains for 6 hours to final cell densities ranging from ABS_{600} ~0.004 to ABS_{600} ~1.0 (Figure 4.4B). Our reasoning was that extracellular L-arabinose would be more depleted at higher densities, but only in the strain with catabolism. For this strain, we observed higher EC_{50} values and lower Hill coefficients with increasing cell densities, paralleling model predictions (Figure 4.4C, Table 4.1). For the strain lacking catabolism, we
observed similar EC$_{50}$ values and Hill coefficients for all cell densities. These findings provide further support for a surprising benefit of sugar catabolism at higher densities and offer a separate form of negative feedback in sugar utilization that may generate more desirable titratable responses. A notable downside to this strategy is that the extent of induction peaks and then decreases over time at higher cell densities (Figure S4.5), potentially complicating the tuning of expression levels.

4.3 Similar trends when linearizing the response to D-xylose

To extend the generality of our insights, we focused on the D-xylose utilization pathway in *E. coli*. Similar to the L-arabinose utilization pathway, the D-xylose utilization pathway encodes a high-affinity transporter and a low-affinity transporter, a transcriptional activator that recognizes D-xylose, and enzymes that shunt D-xylose into the pentose phosphate pathway (Figure S4.1B). Also paralleling the L-arabinose utilization pathway, the D-xylose utilization pathway exhibits a bimodal response when tracking the activity of the *xylA* promoter (Figure 4.5A). We first asked how constitutively expressing the native high-affinity transporter (*xylFGH*) impacts the response. Paralleling the L-arabinose utilization pathway, constitutively expressing the endogenous copy of *xylFGH* resulted in a graded response at all examined D-xylose concentrations (Figure 4.5B). We next deleted the catabolic operon (*xylAB*) to evaluate the impact of removing D-xylose catabolism. Further paralleling the L-arabinose utilization pathway, deletion of the catabolic operon significantly reduced the EC$_{50}$ value (99.8 μM ×½ 2.20 to 0.033 μM ×½ 1.10, p-value = 0.0014) at the cost of an increased Hill coefficient (1.00 ± 0.08 to 1.58 ± 0.24, p-value = 0.019) (Figure 4.5B). Collectively, we were able to generate a linear response to D-xylose through a single alteration and observed similar trade-offs when further deleting sugar catabolism, all matching our observations for the L-arabinose utilization pathway.
Figure 4.4. Effect of cell density in the presence or absence of sugar catabolism. (A) Model predictions for the pathway with a constitutively expressed low-affinity transporter ($T_L = 0.2$) and the deleted high-affinity transporter ($\alpha_H = 0$) when accounting for depletion of extracellular sugar through catabolism. Each simulation was conducted to $\tau = 10$. The different curves reflect the relative volume of the cells to the medium ($v$). Note that all variables were non-dimensionalized as part of the model derivation. See Supporting Information for more details. (B) Growth curves for the $P_{con}$-araE $\Delta$araFGH strain with or without ($\Delta$araBAD) sugar catabolism in defined medium with or without 10 mM L-arabinose. Each value represents the mean of three independent experiments. The S.E.M. for each measurement was smaller than the symbol. (C) Representative dot plots for both strains in log phase grown to the indicated final cell densities. See Figure 4.3 for more information on the dot plots. Each dot plot is representative of at least three experiments conducted from independent colonies. See Table 4.1 for the response metrics that account for the replicate experiments.
4.4 Design rules to engineer sugar utilization pathways for titratable control

Overall, our combination of mathematical modeling and experimental probing of L-arabinose and D-xylose utilization suggest general design rules when co-opting natural sugar utilization pathways as titratable systems. The principal rule is that there is no single, perfect set of alterations within the commonly used set that we investigated. Instead, each alteration comes with specific trade-offs beyond the required number of genomic alterations. Accordingly, the best set of alterations will depend on the demands of the desired titratable system. For instance, systems that rely on expensive inducers would opt for modifications that minimize the EC$_{50}$ value, whereas systems that require fine-tuning would minimize the Hill coefficient. The trade-offs principally applied to the EC$_{50}$ value and the Hill coefficient, whereas alterations did not predictably impact the dynamic range.

The next rule pertains to the need for a uniform response free of bimodality. A uniform response can be most readily achieved by constitutively expressing the high-affinity transporter rather than the low-affinity transporter. Theoretically, sufficient expression of either transporter can eliminate bimodality, although the high-affinity transporter could be expressed at much lower levels. The advantage of this strategy is that lower expression conserves cellular resources and limits the toxicity of overexpressing membrane proteins.

The third rule relates to trade-offs associated with removing one transporter when the other is constitutively expressed. Removing the high-affinity transporter can help eliminate bimodality at the potential cost of a higher EC$_{50}$ value, whereas removing the low-affinity transporter reduces the EC$_{50}$ value at the cost of a sharper response. An alternative would be constitutively expressing the second transporter, which could be explored in future studies.

A final rule pertains to sugar catabolism. Aside from requiring higher sugar concentrations to induce the pathway, sugar catabolism helps reduce the extent of bimodality and linearizes the response. Linearization may be further enhanced when cultures are grown to higher cell densities. Achieving consistent densities between experiments may be challenging and entry into stationary phase may alter pathway activity (Figure S4.5). Furthermore, sugar catabolism may perturb central metabolism as part of metabolic engineering that complicates the results. However, the enhanced linearity would offer finely
tuned control over gene expression that may be essential when precise pathway optimization is needed.

**Figure 4.5.** Linearizing the response to D-xylose. The wild type *E. coli* strain (A), the strain constitutively expressing the high-affinity transporter *xylFGH* (B), and the strain constitutively expressing the high-affinity transporter *xylFGH* and lacking the catabolic operon *xylAB* (C) each harbored the reporter plasmid pUA66-pxylA. The designated strain was back-diluted into M9 minimal medium supplemented with the indicated concentration of D-xylose and grown for 6 hours to ABS$_{600}$ ~0.4 prior to flow cytometry analysis. See Figure 4.3 for more information on the dot plots. Each dot plot is representative of at least three experiments conducted from independent colonies.

The design rules described above were extracted from a simple model and probing of L-arabinose utilization. The rules also applied to the D-xylose utilization pathway, suggesting broader applicability. However, many utilization pathways exhibit variations on the simple pathway, such as anabolic pathways$^{45}$ or pathways that are induced by metabolic intermediates$^{46}$. Further analysis of these pathways will offer insights into the desirability of their natural responses and any required manipulations. Separately, we relied on a more restrained parameter set to probe alterations. Further interrogating parameter space may reveal other parameters and design rules tailored to the behavior of the native pathway. With additional insights, a set of more detailed design principles could be developed. The end result will be reliable means to convert the myriad of sugar utilization pathways in the
microbial world into versatile titratable systems for fundamental genetic studies and for the engineering of microorganisms with optimized metabolic processes.

**Author Information**

Current address for K.B.: Division of Immunobiology and Center for Systems Immunology Cincinnati Children's Hospital Medical Center 231 Albert Sabin Way Cincinnati, OH 45229.

C.L.B. and T.A. conceived of the study, T.A. and K.E.B. performed the experiments, C.L.B. and T.A. designed and interpreted the experiments, K.B. performed and interpreted the modeling, K.B., Y.K., and C.L.B. designed the model, and C.L.B., T.A. and K.B. wrote the manuscript. The authors declare no conflicts of interest.
Table 4.1. Response metrics for each alteration of the L-arabinose utilization pathway in *E. coli*. Listed values are the mean and S.E.M. (dynamic range, δ; Hill coefficient, η) or the geometric mean and geometric S.E.M. (effective concentration to achieve 50% induction, EC<sub>50</sub>) for three independent experiments conducted on separate days. All values reflect cultures grown to ABS<sub>600</sub> ~0.4 unless indicated otherwise. The dynamic range was calculated as the ratio of the maximal to minimal mean fluorescence of the entire population, with autofluorescence subtracted from each value. The large error associated with the dynamic range can be attributed to basal levels approaching autofluorescence of cells lacking the reporter plasmid. Hill coefficient, η. Supporting Information Available: Figures S4.1 – S4.5, Table S4.1 – S4.4, and a detailed description of the simple mathematical model. This material is available free of charge via the Internet at [http://pubs.acs.org](http://pubs.acs.org).

<table>
<thead>
<tr>
<th>Alteration</th>
<th>Bimodality (Y/n)</th>
<th>δ</th>
<th>EC&lt;sub&gt;50&lt;/sub&gt; (μM)</th>
<th>η</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (WT)</td>
<td>Y</td>
<td>958 ± 118</td>
<td>205 ×/÷</td>
<td>0.89 ± 0.19</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>Y</td>
<td>1,023 ± 235</td>
<td>9.8 ×/÷</td>
<td>1.56 ± 0.14</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH</td>
<td>n</td>
<td>290 ± 114</td>
<td>160 ×/÷ 1.47</td>
<td>0.96 ± 0.06</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH ΔaraBAD</td>
<td>n</td>
<td>900 ± 288</td>
<td>0.54 ×/÷ 1.20</td>
<td>1.44 ± 0.07</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH ΔaraE</td>
<td>n</td>
<td>262 ± 17</td>
<td>56 ×/÷ 1.16</td>
<td>1.18 ± 0.03</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH ΔaraE ΔaraBAD</td>
<td>n</td>
<td>1,440 ± 300</td>
<td>0.48 ×/÷ 1.15</td>
<td>1.65 ± 0.03</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE</td>
<td>Y</td>
<td>828 ± 282</td>
<td>284 ×/÷ 1.02</td>
<td>0.75 ± 0.04</td>
</tr>
<tr>
<td>Expression Construct</td>
<td>Y (n)</td>
<td>x/±</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraBAD</td>
<td>1,351 ± 265</td>
<td>0.86 ×/÷</td>
<td>3.27 ± 0.30</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH</td>
<td>737 ± 323</td>
<td>626 ×/÷</td>
<td>0.76 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH ΔaraBAD</td>
<td>728 ± 407</td>
<td>7.5 ×/÷</td>
<td>1.67 ± 0.09</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH (ABS&lt;sub&gt;600&lt;/sub&gt; ~0.004)</td>
<td>517 ± 141</td>
<td>10.1 ×/÷</td>
<td>1.47 ± 0.03</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH ΔaraBAD (ABS&lt;sub&gt;600&lt;/sub&gt; ~0.004)</td>
<td>663 ± 128</td>
<td>7.1 ×/÷</td>
<td>1.86 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH (ABS&lt;sub&gt;600&lt;/sub&gt; ~0.04)</td>
<td>371 ± 91</td>
<td>21.2 ×/÷</td>
<td>1.25 ± 0.10</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH ΔaraBAD (ABS&lt;sub&gt;600&lt;/sub&gt; ~0.04)</td>
<td>512 ± 78</td>
<td>8.6 ×/÷</td>
<td>1.78 ± 0.08</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH (ABS&lt;sub&gt;600&lt;/sub&gt; ~1.0)</td>
<td>1,060 ± 161</td>
<td>2,050 ×/÷</td>
<td>0.73 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH ΔaraBAD (ABS&lt;sub&gt;600&lt;/sub&gt; ~1.0)</td>
<td>2,410 ± 590</td>
<td>7.7 ×/÷</td>
<td>1.64 ± 0.02</td>
<td></td>
</tr>
</tbody>
</table>

**Acknowledgement**

This work was supported by an Undergraduate Research Grant (to K.E.B.), start-up funds from North Carolina State University (to C.L.B.), a mentoring mini-grant from the ADVANCE D-3 program (to C.L.B.), a Doctoral Dissertation Fellowship awarded to K.B., and grants from the National Institutes of Health (GM086865), the National Science Foundation (CBET-0425882, CBET-0644792) and the Minnesota Supercomputing Institute. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575.

**4.5 Methods**

**Bacterial strains and plasmids.** All strains used in this work were derived from *Escherichia coli* K-12 substrain MG1655 and are listed in Table S4.2. With the exception of [ΔxylAB P<sub>xylF</sub>]:[cat P<sub>con</sub>], all genome modifications were achieved by PCR amplifying the *bla* or *cat* resistance cassette from pKD3 and recombining the linear DNA into NM400 through
mini-λ-mediated recombination. The corresponding oligonucleotides are listed in Table S4.4. To insert the constitutive promoters, the promoter sequence (J23110 from the Registry of Standard Biological Parts) was included in one of the primers used to amplify the resistance cassette followed by a second set of primers to add the homology arms for recombination. Successful recombinants were verified by PCR and by sequencing. The genomic locus with the resistance cassette was then transferred to MG1655 by P1 transduction. In the case of ΔaraBAD::cat, the cat cassette was excised with the pCP20 plasmid to allow for the subsequent insertion of the cat cassette as part of the replacement of the promoters for araE or for araFGH. All P1 transductions were verified by PCR. In the case of [ΔxylAB P_xylF]:[cat P_con], the amplified cat cassette was recombineered directly into MG1655 using pDK46.

Once generated, each strain was transformed with the pUA66-ParaB or pUA66-PxylA reporter plasmid (OpenBiosystems) encoding the low-copy sc101 origin-of-replication, the promoter region of the araBAD operon or the xylAB operon, respectively, and a downstream copy of gfpmut2 with a strong ribosome-binding site sequence. The reporter plasmids along with the cloning and recombination plasmids are listed in Table S4.3. Each reporter plasmid offers a balance between maximizing the fluorescence for high sensitivity and minimizing the copy number to limit titration of the transcription regulators AraC and XylR.

**Growth conditions and media.** Strains harboring the reporter plasmid were streaked out from glycerol stocks onto LB plates supplemented with appropriate antibiotics. Single colonies were then inoculated in 2 ml of M9 minimal medium (1X M9 salts, 10 μg/ml thiamine, 2 mM MgSO₄, 0.1 mM CaCl₂ supplemented with 0.4% glycerol, 0.2% casamino acids, and 0.25 μg/ml kanamycin) and grown overnight at 250 RPM and 37°C. The overnight cultures were then back-diluted into 2 ml of the same medium with varying concentrations of the indicated sugar and grown under the same conditions for 6 h to a final ABS₆₀₀ of ~0.4 unless indicated otherwise. Cell density measurements were performed on a Nanodrop 2000c
(Thermo Scientific). For the time-course experiments (Figure S4.3), cultures were grown for the indicated time to the same final $\text{ABS}_{600}$ of $\sim 0.4$.

**Flow cytometry analysis.** Cells were diluted 1:100 in 1X PBS and run on an Accuri C6 flow cytometer (Becton Dickinson) equipped with CFlow plate sampler, a 488 nm laser, and a 530 ± 15 nm bandpass filter. Cells were gated based on forward scatter (FSC-H) and side scatter (SSC-H) using a gate set based on experiments with DRAQ5 dye (Thermo Scientific). A lower cutoff of 11,500 A.U. for FSC-H and 500 A.U. for SSC-H was used to eliminate the appearance of excessive noise. At least 20,000 cells were collected for each sample. See Figure S4.2 for more information on the analysis.

**Curve fitting to extract performance metrics.** The Hill equation ($Y = a \times X^n/(K^n + X^n) + b$, where $a$, $K$, and $n$ are fit constants and $b$ is the fluorescence in the absence of the inducing sugar) was used to fit the mean fluorescence values ($Y$) across the different applied concentrations of applied sugar ($X$). Curve fitting was performed using the least-squares approach with the natural log of the measured and predicted mean fluorescence values. Of the fit values, $n$ is the Hill coefficient ($\eta$), $K$ is the $\text{EC}_{50}$ value, and $(a + b - \text{AF})/(b - \text{AF})$ is the dynamic range ($\delta$) (where $\text{AF}$ is autofluorescence of the cells harboring the pUA66 plasmid).

**Mathematical modeling.** All simulations were conducted in MATLAB. The ordinary differential equations were integrated using the Euler method. The stable and unstable steady-states were calculated by implementing the arc-length continuation method. The details of the model can be found in Supporting Information, including its derivation and the selected values for the model parameters.

**Statistical analyses.** P-values were calculated using the student t-test with unequal variance when comparing all parameter metrics. The test used one tail for expected changes between
metrics (e.g. an increase in the Hill coefficient) and two tails when no changes were expected. EC$_{50}$ values were assumed to follow a log-normal distribution.
References


Chapter 4

Trade-offs in engineering sugar utilization pathways for titratable control
(Supplementary Information)

*This work has been published in ACS Synthetic Biology. Contributions to this work: Kelsey Boykin, Konstantinos Biliouris, Yiannis Kaznessis*
**Description of simple model**

We developed a simple deterministic model comprising six ordinary differential equations to capture the dynamical behavior of utilization pathways that catabolize the inducing sugar. The model accounts for the regulator (R), the extracellular (S₀) and intracellular (S) sugar, two transporters (Tₜ, Tₜ₊), and one enzyme (E). Binding of S to R generates an active regulator (R) that upregulates the expression of the transporters and the enzyme following the Hill equation. The transporters import sugar into the cell while the enzyme irreversibly degrades the sugar. The model also allows for the breakdown of S₀ through sugar catabolism. The volume ratio of cells and medium converts between the amounts of intracellular and extracellular sugar. The following equations describe the amount of R (eq. 1), S₀ (eq. 2) and S (eq. 3), Tₜ and Tₜ₊ (eq. 4, 5), and E (eq. 6). The equations of the model follow mass-action and Michaelis-Menten kinetics, similar to previously published studies describing sugar utilization pathway.¹,²

\[
R = \frac{S}{k_0 + S}
\]

\[
\frac{dS_0}{dt} = \frac{S}{k_3 + S} \frac{E}{V_M} V_C
\]

\[
\frac{dS}{dt} = \frac{S_0}{k_1 + S_0} T_L + \frac{S_0}{k_2 + S_0} T_H + \frac{S}{k_5 + S} E
\]

\[
\frac{dT_L}{dt} = \frac{b_{T_L}}{\text{basal expression of low-affinity transporter}} + \frac{R^n}{k_4 + R^n} \frac{d T_L}{\text{dilation of low-affinity transporter}}
\]

\[
\frac{dT_H}{dt} = \frac{b_{T_H}}{\text{basal expression of high-affinity transporter}} + \frac{R^n}{k_5 + R^n} \frac{d T_H}{\text{dilation of high-affinity transporter}}
\]

\[
\frac{dE}{dt} = \frac{b_E}{\text{basal expression of enzyme}} + \frac{R^n}{k_6 + R^n} \frac{d E}{\text{dilation of enzyme}}
\]
The model does not explicitly account for DNA looping\(^1,2\) or auto-repression of the regulator\(^3\); instead, the model assumes that the total amount of regulator is constant and DNA looping and regulator auto-repression were indirectly incorporated by manipulating the Hill coefficient in the expression terms. The model also neglects differences in growth rate due to the small (~20%) difference in growth rate in the presence or absence of L-arabinose (data not shown). Each parameter in the model is described below.

**Table S4.1 Parameters of the model**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description (Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Concentration of active regulator (M)</td>
</tr>
<tr>
<td>R(_t)</td>
<td>Concentration of total regulator (M)</td>
</tr>
<tr>
<td>S</td>
<td>Intracellular concentration of sugar (M)</td>
</tr>
<tr>
<td>S(_0)</td>
<td>Extracellular concentration of sugar (M)</td>
</tr>
<tr>
<td>t</td>
<td>Time (s)</td>
</tr>
<tr>
<td>V(_C)</td>
<td>Volume of the cell (M)</td>
</tr>
<tr>
<td>V(_M)</td>
<td>Volume of the medium (M)</td>
</tr>
<tr>
<td>T(_L)</td>
<td>Low-affinity transporter concentration (M)</td>
</tr>
<tr>
<td>T(_H)</td>
<td>High-affinity transporter concentration (M)</td>
</tr>
<tr>
<td>E</td>
<td>Enzyme concentration (M)</td>
</tr>
<tr>
<td>k(_0)</td>
<td>Half maximum regulator concentration (M)</td>
</tr>
<tr>
<td>k(_1)</td>
<td>Half-maximum concentration of sugar (M)</td>
</tr>
<tr>
<td>k(_2)</td>
<td>Half-maximum concentration of sugar (M)</td>
</tr>
<tr>
<td>k(_3)</td>
<td>Half-maximum concentration of sugar (M)</td>
</tr>
<tr>
<td>k(_4)</td>
<td>Half-maximum concentration of regulator (M)</td>
</tr>
<tr>
<td>k(_5)</td>
<td>Half-maximum concentration of regulator (M)</td>
</tr>
<tr>
<td>k(_6)</td>
<td>Half-maximum concentration of regulator (M)</td>
</tr>
<tr>
<td>b(_{TL})</td>
<td>Basal expression of first transporter (M/s)</td>
</tr>
<tr>
<td>b(_{TH})</td>
<td>Basal expression of second transporter (M/s)</td>
</tr>
<tr>
<td>b(_E)</td>
<td>Basal expression of enzyme (M/s)</td>
</tr>
<tr>
<td>α(_1)</td>
<td>Maximum rate of sugar import from low-affinity transporter (1/s)</td>
</tr>
<tr>
<td>α(_2)</td>
<td>Maximum rate of sugar import from high-affinity transporter (1/s)</td>
</tr>
<tr>
<td>α(_3)</td>
<td>Maximum rate of sugar catabolism (M/s)</td>
</tr>
<tr>
<td>α(_4)</td>
<td>Maximum rate of low-affinity transporter production (M/s)</td>
</tr>
<tr>
<td>α(_5)</td>
<td>Maximum rate of high-affinity transporter production (M/s)</td>
</tr>
<tr>
<td>α(_6)</td>
<td>Maximum rate of enzyme production (M/s)</td>
</tr>
<tr>
<td>d</td>
<td>Dilution rate due to cell division (1/s)</td>
</tr>
<tr>
<td>n</td>
<td>Hill coefficient</td>
</tr>
</tbody>
</table>
To reduce the total number of parameters, we first non-dimensionalized the variables as shown on the next page.

**Table S4.2 Dimensionless parameters**

<table>
<thead>
<tr>
<th>Dimensionless variable</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R'$</td>
<td>$R/R_t$</td>
</tr>
<tr>
<td>$S'$</td>
<td>$S/k_0$</td>
</tr>
<tr>
<td>$S_0'$</td>
<td>$S_0/k_1$</td>
</tr>
<tr>
<td>$T_{L'}$</td>
<td>$(d/\alpha_4)\cdot T_1$</td>
</tr>
<tr>
<td>$T_{H'}$</td>
<td>$(d/\alpha_5)\cdot T_2$</td>
</tr>
<tr>
<td>$E'$</td>
<td>$(d/\alpha_6)\cdot E$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>$t-d$</td>
</tr>
</tbody>
</table>

Non-dimensionalizing the variables resulted in the following equations:

$$R' = \frac{S'}{1+S'}$$  \hspace{1cm} (1b)

$$\frac{dS_0'}{d\tau} = \frac{\alpha_5\alpha_6}{k_0d^2} \frac{S'}{k_3 + S'}E' \frac{k_0}{k_1} \frac{V_c}{V_M}$$  \hspace{1cm} (2b)

$$\frac{dS'}{d\tau} = \frac{\alpha_5\alpha_4}{k_0d^2} \frac{S_0'}{1+S_0'}T_{L'} + \frac{\alpha_5\alpha_6}{k_0d^2} \frac{k_3}{k_1} \frac{k_2}{k_3 + S'}S_0' E' - S'$$  \hspace{1cm} (3b)

$$\frac{dT_{L'}}{d\tau} = b_{L'} + \frac{(R')^n}{\alpha_4 \left( \frac{k_4}{R_t} \right)^n} - T_{L'}$$  \hspace{1cm} (4b)

$$\frac{dT_{H'}}{d\tau} = b_{H'} + \frac{(R')^n}{\alpha_5 \left( \frac{k_5}{R_t} \right)^n} - T_{H'}$$  \hspace{1cm} (5b)

$$\frac{dE'}{d\tau} = b_E + \frac{(R')^n}{\alpha_6 \left( \frac{k_6}{R_t} \right)^n} - E'$$  \hspace{1cm} (6b)
We next assigned single constants to the lumped parameters to generate the following set of equations:

\[
R' = \frac{S'}{1 + S'} \quad \text{(1c)}
\]

\[
\frac{dS'_0}{d\tau} = -\alpha_E \frac{S'}{K_E + S'} E' \nu \quad \text{(2c)}
\]

\[
\frac{dS'}{d\tau} = \alpha_L \frac{S'_0}{1 + S'_0} - T'_1 + \alpha_H \frac{S'_0}{K_H + S'_0} - T'_2 - \alpha_E \frac{S'}{K_E + S'} E' - S' \quad \text{(3c)}
\]

\[
\frac{dT'_L}{d\tau} = b'_L + \frac{(R')^n}{K_{RL}^n + (R')^n} - T'_L \quad \text{(4c)}
\]

\[
\frac{dT'_H}{d\tau} = b'_H + \frac{(R')^n}{K_{RH}^n + (R')^n} - T'_H \quad \text{(5c)}
\]

\[
\frac{dE'}{d\tau} = b'_E + \frac{(R')^n}{K_{RE}^n + (R')^n} - E' \quad \text{(6c)}
\]

The final constants along with the equivalent lumped parameters are shown in the table on the next page. We selected values for each constant to capture the general behavior of the L-arabinose utilization pathway: a high-affinity/low-capacity transporter expressed at lower sugar concentrations and a low-affinity/high-capacity transporter expressed at higher sugar concentrations \((\alpha_L > \alpha_H, K_L < K_H, K_{RL} < K_{RH})\), low basal expression \((b'_L, b'_H, b'_E < 1)\), proteins and sugar that are diluted through cell division, and sufficient protein expression and activity \((\alpha_E, K_E)\) to yield a bistable response.
Table S4.3 Values of the dimensionless parameters

<table>
<thead>
<tr>
<th>Formula</th>
<th>Dimensionless parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{k_0 d^2}$</td>
<td>$a_L$</td>
<td>300</td>
</tr>
<tr>
<td>$\frac{k_2}{k_0}$</td>
<td>$K_L$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\frac{2}{k_0 d^2}$</td>
<td>$a_H$</td>
<td>100</td>
</tr>
<tr>
<td>$\frac{k_3}{k_0}$</td>
<td>$K_E$</td>
<td>0.01</td>
</tr>
<tr>
<td>$\frac{3}{k_0 d^2}$</td>
<td>$a_E$</td>
<td>0.6</td>
</tr>
<tr>
<td>$\nu$</td>
<td>$\frac{k_0 V_C}{k_1 V_M}$</td>
<td>$0 - 0.01$</td>
</tr>
<tr>
<td>$\frac{k_4}{R_t}$</td>
<td>$K_{RL}$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\frac{k_5}{R_t}$</td>
<td>$K_{RH}$</td>
<td>0.005</td>
</tr>
<tr>
<td>$\frac{k_6}{R_t}$</td>
<td>$K_{RE}$</td>
<td>0.02</td>
</tr>
<tr>
<td>$\frac{b_1}{R_t}$</td>
<td>$b_L'$</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\frac{b_2}{R_t}$</td>
<td>$b_H'$</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\frac{b_3}{R_t}$</td>
<td>$b_E'$</td>
<td>0.0002</td>
</tr>
<tr>
<td>$n$</td>
<td>$n$</td>
<td>2</td>
</tr>
</tbody>
</table>
These values were used in all MATLAB simulations unless indicated otherwise. The simulation results in Figure 4.2 neglected sugar breakdown, which involved removing the differential equation for extracellular sugar (eq. 2c). To capture constitutive expression of either transporter, we replaced the corresponding expression term \( \frac{R_2^2}{K_2 + R_2^2} \) with the indicated value. To delete either transporter or the enzyme, we set the maximal activity term equal to zero (e.g. \( \alpha_L = 0 \)). As part of the simulations that neglected sugar breakdown, we integrated the equations using Euler’s method and calculated the steady-state curves using the arc-length continuation method. When accounting for sugar breakdown, we integrated the equations to \( \tau = 10 \), which was sufficient to reach steady-state in the absence of breakdown. Altering the final value of \( \tau \) did not change the qualitative impact of sugar breakdown.
Supporting figures

**Figure S4.1.** The L-arabinose and D-xylose utilization pathways in *E. coli.* (A) The pathway is composed of one regulator (*araC*), one high-affinity transporter (*araFGH*), one low-affinity transporter (*araE*), one putative transporter (*araJ*), and three enzymes (*araBAD*) that shunt L-arabinose into the pentose phosphate pathway (PPP). The AraC protein activates transcription at each promoter when bound to L-arabinose. Although the AraJ protein resembles a transporter, previous deletion of *araJ* had no measurable impact on the behavior of the pathway. (B) The D-xylose utilization pathway is structured similarly to the L-arabinose utilization pathway, with a high-affinity transporter (*xylFGH*), a low-affinity transporter (*xylE*), two enzymes (*xylAB*) that shunt D-xylose into the pentose phosphate pathway, and a single regulator (*xylR*) responsible for activating the operons in the presence of D-xylose.
Figure S4.2. Overview of flow cytometry analysis. Cultures were diluted 1:100 in 1X PBS and run on a BD Accuri C6 personal flow cytometer. Cells were gated based on forward scatter (FSC-H) and side scatter (SSC-H). The gate (red) was selected to capture cells and no “noisy” events based on initial runs with DRAQ5 dye. Once gated, the mean fluorescence was extracted for unimodal distributions, while the mean fluorescence and relative number of events for each subpopulation were extracted for bimodal distributions.
Figure S4.3. Impact of incubation time on the response to L-arabinose. Cultures of the WT strain (CB26) were grown in varying L-arabinose concentrations for the indicated amount of time to mid-log phase (ABS$_{600}$ ~0.4) and then subjected to flow cytometry analysis. The concentration to achieve 50% induced cells was estimated by extracting the percentage of the population that was induced and fitting these values to a Hill equation. Note that this approach was selected over calculating the EC$_{50}$ value because the response curve did not saturate at the maximal L-arabinose concentration that we used (10 mM). The reported concentration values represent the geometric mean and S.E.M. from at least three independent experiments. We selected an incubation time of 6 hours for all other experiments because increasing the incubation time had no obvious effect on the response.
Figure S4.4. Predicted response when only one transporter is present. Model predictions reflect the base model in Figure 4.2 when the maximal activity of either transporter ($\alpha_H$, $\alpha_L$) is set equal to zero. Both transporters are sufficient for full induction of the pathway, where much lower concentrations of the extracellular sugar are needed for the high-affinity transporter than for the low-affinity transporter. Note that all variables were non-dimensionalized as part of the model derivation.
Figure S4.5. Response dynamics in the presence or absence of L-arabinose catabolism. The $P_{\text{con-araE ΔaraFGH}}$ and $P_{\text{con-araE ΔaraFGH ΔaraBAD}}$ strains harboring the pUA66-ParaB plasmid were grown in M9 minimal medium with 292 μM or 8 μM of L-arabinose, respectively. These L-arabinose concentrations reflect the approximate EC$_{50}$ value for each modified pathway. The ABS$_{600}$ (A) and single-cell fluorescence (B) were measured at the indicated times. The values represent the mean and S.E.M. from experiments conducted with three independent colonies. The error bars are smaller than each symbol at all time points. The ABS$_{600}$ was too low to accurately measure prior to 3 hours.
### Supporting Tables

**Table S4.4.** Measured growth rates for each strain. Measurements were performed in M9 glycerol with 0.2% casamino acids and either no L-arabinose (- ara) or 10 mM L-arabinose (+ ara) during exponential growth. \(^a\) Growth rates reflect the mean and S.E. from the measurement of three independent colonies.

<table>
<thead>
<tr>
<th>Alteration</th>
<th>Growth rate(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- ara</td>
</tr>
<tr>
<td>None (WT)</td>
<td>38.6 ± 1.2</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>38.9 ± 0.2</td>
</tr>
<tr>
<td>(P_{\text{con}})-araFGH</td>
<td>38.7 ± 0.8</td>
</tr>
<tr>
<td>(P_{\text{con}})-araFGH ΔaraBAD</td>
<td>35.7 ± 0.8</td>
</tr>
<tr>
<td>(P_{\text{con}})-araFGH ΔaraE</td>
<td>32.2 ± 1.2</td>
</tr>
<tr>
<td>(P_{\text{con}})-araFGH ΔaraE ΔaraBAD</td>
<td>36.9 ± 1.4</td>
</tr>
<tr>
<td>(P_{\text{con}})-araE</td>
<td>41.5 ± 2.5</td>
</tr>
<tr>
<td>(P_{\text{con}})-araE ΔaraBAD</td>
<td>37.9 ± 1.2</td>
</tr>
<tr>
<td>(P_{\text{con}})-araE ΔaraFGH</td>
<td>34.5 ± 0.5</td>
</tr>
<tr>
<td>(P_{\text{con}})-araE ΔaraFGH ΔaraBAD</td>
<td>37.7 ± 0.5</td>
</tr>
</tbody>
</table>
Table S4.5. Strains used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Genotype</th>
<th>Stock #</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td><em>Escherichia coli</em> K-12 substrain</td>
<td>CB168</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>MG1655 ΔaraBAD::cat</td>
<td>CB225</td>
</tr>
<tr>
<td>NM400</td>
<td>MG1655 mini-λ-cat</td>
<td>CB25</td>
</tr>
<tr>
<td>NM500</td>
<td>MG1655 mini-λ-tetA</td>
<td>CB26</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE</td>
<td>MG1655 P&lt;sub&gt;araE&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;]</td>
<td>CB325</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraBAD</td>
<td>MG1655 P&lt;sub&gt;araE&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;] ΔaraBAD</td>
<td>CB326</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH</td>
<td>MG1655 P&lt;sub&gt;araE&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;] ΔaraFGH::bla</td>
<td>CB327</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araE ΔaraFGH</td>
<td>MG1655 P&lt;sub&gt;araE&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;] ΔaraBAD</td>
<td>CB328</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>ΔaraFGH::bla</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH</td>
<td>MG1655 P&lt;sub&gt;araF&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;]</td>
<td>CB329</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH</td>
<td>MG1655 P&lt;sub&gt;araF&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;] ΔaraBAD</td>
<td>CB330</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>ΔaraE::bla</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH ΔaraE</td>
<td>MG1655 P&lt;sub&gt;araF&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;] ΔaraE::bla</td>
<td>CB331</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-araFGH ΔaraE</td>
<td>MG1655 P&lt;sub&gt;araF&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;] ΔaraBAD</td>
<td>CB332</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>ΔaraE::bla</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-xylFGH</td>
<td>MG1655 P&lt;sub&gt;xylF&lt;/sub&gt;::[cat P&lt;sub&gt;con&lt;/sub&gt;]</td>
<td>CB342</td>
</tr>
<tr>
<td>P&lt;sub&gt;con&lt;/sub&gt;-xylFGH ΔxylAB</td>
<td>MG1655 [ΔxylAB P&lt;sub&gt;xylF&lt;/sub&gt;][cat P&lt;sub&gt;con&lt;/sub&gt;]</td>
<td>CB343</td>
</tr>
</tbody>
</table>
**Table S4.6.** Plasmids used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Source</th>
<th>Stock #</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUA66</td>
<td>Low-copy plasmid (pSC101 ori) with MCS upstream of <em>gfp</em></td>
<td>OpenBiosystems</td>
<td>pCB198</td>
</tr>
<tr>
<td>pUA66ParaB</td>
<td>pUA66 with the <em>araB</em> promoter</td>
<td>OpenBiosystems</td>
<td>pCB208</td>
</tr>
<tr>
<td>pUA66PxyIA</td>
<td>pUA66 with the <em>xylA</em> promoter</td>
<td>OpenBiosystems</td>
<td>pCB289</td>
</tr>
<tr>
<td></td>
<td>Plasmid encoding <em>cat</em> cassette</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pKD3</td>
<td>flanked by FRT sites and a <em>bla</em> cassette.</td>
<td>Ref. 5</td>
<td>pCB333</td>
</tr>
<tr>
<td></td>
<td>L-arabinose-inducible expression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pKD46</td>
<td>of <em>λ</em>-red genes on a plasmid with a heat-sensitive origin-of-replication</td>
<td>Ref. 5</td>
<td>pCB13</td>
</tr>
<tr>
<td></td>
<td>Heat-inducible expression of FRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCP20</td>
<td>on a plasmid with a heat-sensitive origin-of-replication</td>
<td>Ref. 6</td>
<td>pCB44</td>
</tr>
</tbody>
</table>
**Table S4.7.** Oligonucleotides used in this study. Bases shown in red are part of the synthetic promoter (J23110) are shown red. Bolded bases are the homology regions for recombination into MG1655.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>del-araBAD.fwd</td>
<td>TCCATACCCGTTTTTTTTGGATGGAGTGAAGACGACATG</td>
</tr>
<tr>
<td></td>
<td>GTCCATATGAAATATCTCCTCTTTAG</td>
</tr>
<tr>
<td>del-araBAD.rev</td>
<td>GTTTGCTGCAATATCCGGTAACTCGGCGCTAACTGA</td>
</tr>
<tr>
<td></td>
<td>CGGCAGGTAGGCTGGAGCTGCTTT</td>
</tr>
<tr>
<td>ParaE.fwd</td>
<td>TCTGCTGTTAATATTAGGTGTTAAATAATATCCTCAAT</td>
</tr>
<tr>
<td></td>
<td>ATATTCAAGCTGGAGCTGCTTT</td>
</tr>
<tr>
<td>ParaF.fwd</td>
<td>CCATTTCAAAAACTCAACGCGCCAGTAATGCGCCCTA</td>
</tr>
<tr>
<td></td>
<td>TTGAAGCTGGAGCTGCTTT</td>
</tr>
<tr>
<td>J23110.rev</td>
<td>GCTAGCATTGTACCTAGGACTGAGCTACCGTAACAC</td>
</tr>
<tr>
<td></td>
<td>ATATGAAATATCCTCCTTTAG</td>
</tr>
<tr>
<td>ParaE-J23110.rev</td>
<td>GATAGTAAACCATTTTTTTCTGCCACGAGAGTAAG</td>
</tr>
<tr>
<td>J23110.rev</td>
<td>ACGTAGATCATTGTACCTAGGACTG</td>
</tr>
<tr>
<td>ParaF-J23110.rev</td>
<td>AAACATTACCTTTAAACTAAAAAGATAAGTGAATGT</td>
</tr>
<tr>
<td>J23110.rev</td>
<td>GTGCTAGCATTGTACCTAGGACTG</td>
</tr>
<tr>
<td>del-araE.fwd</td>
<td>TGTTACAGTATTTTTTCAACTATGTCTTACTCTCTGCT</td>
</tr>
<tr>
<td>del-araE.rev</td>
<td>GACGCGATATTTCTCATACTTTCTCGCTGCCATCAG</td>
</tr>
<tr>
<td></td>
<td>TTTTACCAATGCTTTAATCAGTGT</td>
</tr>
</tbody>
</table>
Table S4.7. Continued

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>del- araFGH.fwd</td>
<td>TCATTCGTTTTTGCCCTACACAAACGACACCTAAAG</td>
</tr>
<tr>
<td>del- araFGH.rev</td>
<td>CTGGTGTAGCCTGGGCAGTCTCCCTA</td>
</tr>
<tr>
<td>del-</td>
<td>GACAGTGCCTTTGTCTTTTGCTTGAACGGGTGAA</td>
</tr>
<tr>
<td>PxylA.fwd</td>
<td>TATCGAGGCTCCCTTCCTTCACGGGTGTAGGCGCTT</td>
</tr>
<tr>
<td>PxylA.rev</td>
<td>TGTATATCCTCTCTTAAATCTAGATTACATACGCTTA CCAACACCA</td>
</tr>
<tr>
<td>PxylF.fwd</td>
<td>GCAATAAGTCAATTCGCAACAAAAAAGTAAGATCTC GGTCATAGTAGGCTGAGCTT</td>
</tr>
<tr>
<td>PxylF- J23110.rev</td>
<td>GGACAGTTTTAATAAGTAACACATCCACCCGATAAAC</td>
</tr>
<tr>
<td>del-</td>
<td>GGCAAAATTCATGAGCGATTTCTTCTGGATTGC</td>
</tr>
<tr>
<td>xylAB.rev</td>
<td>GGCATCGTAGGCTGGAGCTGCTT</td>
</tr>
</tbody>
</table>

Supporting References


Chapter 5

Impact of residual inducer in titratable expression systems

*This work will be submitted to Biotechnology Journal. Contributions to this work: Michelle Luo
Abstract
Inducible expression systems are widely employed for the titratable control of gene expression, yet some systems can respond to inducing molecules inadvertently present in the growth medium or synthesized by the host cells. Here, we explored the impact of these residual inducers on the response properties of inducible systems. Using mathematical modeling, we found that the presence of residual inducer shrinks the dynamic range and causes the apparent Hill coefficient to converge to one. We also found that inducing systems were more sensitive than repressing systems to the presence of residual inducer and the effects of residual inducer on the response parameters are heavily influenced by the Hill coefficient of the original response. Experimental interrogation of common inducible systems based on an L-arabinose inducible promoter or a thiamine pyrophosphate-responsive riboswitch in the bacterium *Escherichia coli* confirmed the predicted trends. Interrogation of an “all-or-none” inducible system showed that the cells are sensitive to residual inducer when they exhibit bimodal response but not when they become fully induced. Our findings indicate that residual inducer—whether a component of the growth medium or manufactured by the cells—alters the quantitative response properties of inducible systems, impacting their utility for scientific discovery and pathway engineering.
5.1 Introduction

Separately, multiple sugar utilization pathways in bacteria (e.g. the D- Inducible expression systems have proven to be invaluable tools for probing gene function and optimizing the levels of pathway components. These systems traditionally rely on the addition of a small-molecule inducer that enters the cell either passively or via active transport. The inducer then interacts with a signaling cascade or an intracellular sensory regulator, thereby modulating gene expression. Classically, transcriptional sensors are employed as inducible systems that activate or repress transcription in the presence of the inducer (Brautaset et al., 2009; Mizuguchi and Hayakawa, 2002). Common examples include the P_{BAD} promoter in *E. coli* that is transcriptionally activated when the AraC regulator bound to the inducer L-arabinose recruits RNA polymerase (Schleif, 2010) or the TetOn system in higher eukaryotes in which the engineered rtTA is fused to the VP16 activator domain, resulting in DNA binding and polymerase recruitment in the presence of the small molecule doxycycline (Meyer-Ficca et al., 2004). Aside from these classical systems, riboswitches and other RNA-based devices have become alternative means for conditional control of gene expression. Riboswitches undergo conformational changes when bound to their cognate inducer, resulting in modulation of the transcription or translation of downstream genes. Regardless of the exact mechanism of action, the simple addition of a small-molecule inducer to the growth medium or environment has facilitated the facile and finely tuned control of gene expression.

With the widespread use of these systems have come reports of inducer being inadvertently present in the culture medium or being manufactured by the cells—what we term ‘residual inducer.’ For instance, carryover of tetracycline in animal feed was present in fetal bovine serum, thereby impacting inducible expression with the TetOn system (Chopra and Roberts, 2001)galactose and N-acetylglucosamine pathways in *E. coli* can synthesize their inducing sugar, offering a separate source of inducer when employing pathway promoters as titratable systems (Afroz et al., 2014b; Vogler et al., 1989; Weickert and Adhya, 1993). While these complications might lead some to discard these inducible systems or reformulate the growth medium, the underlying question is how these sources of inducer impact the apparent relationship between exogenously added inducer and expression levels.
of regulated genes. Understanding this relationship is particularly important for the fine-tuning of gene expression in quantitative genetic studies as well as the construction of gene circuits that are highly sensitive to component parameters. Here, we employed mathematical modeling and experimental interrogation of inducible systems in *E. coli* to determine how residual inducer impacts the observed response properties. We found that the impact of residual inducer on the response parameters depends primarily on the Hill coefficient of the original response (\( \eta \)). For \( \eta \) greater than 1, the residual inducer imparts two concomitant effects on titratable systems: linearizing the response while shrinking the dynamic range. These effects reflect a trade-off, as titratable control typically requires linear responses with large dynamic ranges. Interestingly, residual inducer causes the apparent Hill coefficients (\( \eta' \)) to converge to 1 irrespective of the original values, implying that residual inducer sharpens the responses with \( \eta \) less than 1. Compared to the activating system, the repressing system is less susceptible to residual inducer concentrations. The impact of residual inducer on \( EC_{50} \) values is complex. It exhibits a monotonic response for \( \eta > 1 \) but an increasing trend for \( \eta < 1 \). In contrast, residual inducer exposed to non-titratable or “all-or-none” systems increased the apparent sensitivity to inducer until the sufficient residual inducer fully induced the system. Our results reveal how residual inducer impacts the quantitative properties of inducible systems, providing insights into how the presence of the inducer can be managed when using these systems for fundamental genetic studies and pathway optimization.

### 5.2 Materials and methods

**Bacterial strains and plasmids.** The *E. coli* strains and the pUA66-ParaB reporter plasmid used for the modified L-arabinose utilization pathway were previously reported in (Afroz et al., 2014a). To construct the pUA66-thiC reporter plasmid, the pUA66 plasmid was purified and linearized with BamHI/XhoI. The plasmid backbone was PCR amplified in two parts using the primers sc101.fwd and s101.rev as well as the primers pUA66.fwd and pUA66.rev. The portion of the *thiC* riboswitch and the first 14 codons of *thiC* was amplified from genomic DNA purified from *E. coli* K-12 substrain MG1655 using thiC.fwd and thiC.rev, each having a 5' overhang sequence that would pair with an end of the amplified halves of
pUA66. The synthetic promoter (BBa_J23119 from the Registry of Standard Biological Parts) was encoded in the 5' overhangs of thiC.fwd and pUA66.rev. Gibson assembly was used to assemble the set of three linear pieces of DNA into a single plasmid (Gibson, 2011). Successful recombinants were verified by PCR and by Sanger sequencing.

**Growth conditions and media.** Strains harboring the reporter plasmid were streaked out from freezer stocks into LB plates supplemented with kanamycin and the single colonies were then inoculated in 2 mL of M9 minimal medium (1X M9 salts, 10 μg/mL thiamine, 2 mM MgSO₄, 0.1 mM CaCl₂ supplemented with 0.4% glycerol, 0.2% casamino acids, and 0.25 μg/mL kanamycin) and grown overnight at 250 rpm and 37 °C. The overnight cultures were then back-diluted into 2 mL of the media containing residual inducer with varying concentration of the indicated inducer. The back-diluted cultures were then grown for 6 h under the same conditions to a final ABS₆₀₀ of ~0.4.

**Flow cytometry analysis in E. coli.** Accuri C6 flow cytometer (Becton Dickinson) equipped with CFlow plate sampler, a 488 nm laser, and a 530 ± 15 nm bandpass filter was used for the GFP measurements in E. coli similar to previous work (Afroz et al., 2014a). Cells were diluted 1:100 in 1× PBS before running on the flow cytometer. To eliminate noise, cells were gated based on forward scatter (FSC-H) and side scatter (SSC-H) using a gate set based on experiments with DRAQ5 dye (Thermo Scientific). A lower cutoff of 12,000 au for FSC-H and 500 au for SSC-H was used. At least 20,000 gated events were collected for each sample. The associated dot plots were generated as described previously (Afroz et al., 2014a, 2014b).

**Curve fitting to extract performance metrics.** Autofluorescence of cells harboring pUA66 were deducted from the fluorescence readings. For the activator system, The Hill equation given by Eq. 1

\[ y = \frac{A(x)^\eta}{K^\eta + (x)^\eta} + B \]  

(Eq. 1)
where A, K, and η are fit constants and B is the fluorescence in the absence of the inducing sugar) was used to fit the mean fluorescence values (y) across the different applied concentrations of applied inducer (x).

Curve fitting was performed using the least-squares approach with the natural log of the measured and predicted mean fluorescence values. Of the fit values, n is the Hill coefficient (η), K is the EC$_{50}$ value, and (A + B)/(B) is the dynamic range (δ). An equivalent approach was used to extract the parameter values for the repressing system using a Hill equation (Eq. 2)

$$y = \frac{A}{1 + (\frac{x}{K})^\eta} + B$$  
(Eq.2)

**Mathematical modeling.** Using the Hill equations for activating systems (Eq. 1) or repressing systems (Eq. 2), we modeled the effects of residual inducer on the response parameters:

$$y = \frac{A(x+c)^\eta}{K^\eta + (x+c)^\eta} + B$$  
(Eq.3)

$$y = \frac{A}{1 + (\frac{x+c}{K})^\eta} + B$$  
(Eq.4)

where y is the gene expression output, x is the concentration of applied inducer, c is the concentration of residual inducer, K is the half-maximal inducer concentration (EC$_{50}$), η is the Hill coefficient, and B is basal level. For the activating system, A’=A+B-y(c) where A and B are original values of A and B respectively. For the repressing system, A’=y(c)-B and B’=B The natural log of the expression levels with residual inducer were then fit to the natural log of Equations 5 and 6 using the least-squares method to determine the apparent constants (K, η) that would approximate the same response with no basal induction.
\[ y = \frac{A'(x)\eta'}{(K')\eta' + (x)\eta'} + B' \]  \hspace{1cm} (Eq. 5)

\[ y = \frac{A'}{1 + \left(\frac{x}{K}\right)^\eta'} + B' \]  \hspace{1cm} (Eq. 6)

The dynamic range (δ) was subsequently calculated according to the following (Eq. 7):

\[ \delta = \frac{A' + B'}{B'} \]  \hspace{1cm} (Eq. 7)

5.3 Results

5.3a Mathematical modeling to assess the effects of residual inducer for an inducing system in E. coli

Regulation by gene activation is commonly employed in *Escherichia coli*. We thereby first asked how residual inducer would quantitatively impact the apparent response curve of an activating system. Figure S5.1 illustrates the parameters we used to evaluate the apparent response and the impacts of the residual inducer on the parameter values. The parameters used are Hill coefficient (\(\eta\)) given by the slope of the response; EC\(_{50}\), which is the inducer concentration that yields the average of the maximal and minimal expression levels and dynamic range (\(\delta\))—calculated as the ratio of maximal to minimal gene expression levels.

We considered a simple response curve captured by a modified Hill Equation (Eq. 3) in which inducer comes from two sources: inducer added to the medium (\(x\)) and residual inducer already present in the medium (\(c\))—whether inadvertently present or manufactured by the cells. The traditional Hill equation was then fit to the resulting curves assuming that the inducer added to the medium was the only inducer known to be present in the medium. The resulting impact on the apparent response curves for varying Hill coefficients, (Figure 5.1A (top panel)) a EC\(_{50}\) and dynamic range (Fig. S5.2) are found using mathematical modeling.

Based on the fit values, we observed trends for the activating system that depended strongly on the original Hill coefficient values. First, for \(\eta > 1\), the residual inducer reduced
the apparent Hill coefficient (η’). The less steep curve comes from stretching the response curve from the residual inducer concentration to a value of zero, giving the curve a more graded appearance (Fig. S5.1). Also, the residual inducer had a non-monotonic impact on the apparent half-maximal inducer concentration (EC50) for η>1. Interestingly, the residual inducer sharpens the responses with η<1 and the apparent Hill coefficient converged on a value of one for all η values. Furthermore, with η≤ 1, EC50 values exhibit an upward trend with increase in residual inducer concentrations. For all η values, residual inducer greatly impaired the dynamic range. This anticipated effect can be attributed to residual inducer elevating basal expression, leaving only a portion of the response curve for full induction. Sufficient residual inducer maximally induced expression, thereby eliminating any dynamic range.

As mentioned earlier, in addition to investigating the effects of original Hill coefficient values on response parameters for different residual inducer concentrations, we examined how original dynamic range or EC50 value affects the response parameters (Fig. S5.2) The responses were similar irrespective of the original EC50 or dynamic range values. These observations imply that the response parameter values for different residual inducer concentrations depend mainly on the original Hill coefficient value η.
Figure 5.1. Mathematical modeling of residual inducer on titratable systems. Left and right columns are for inducing and repressing systems, respectively. (A) Residual inducer stretches the response curve to lower applied inducer concentrations. The response curve reflects the relationship between the inducer concentration and gene induction. (B) Predicted effect of residual inducer on the apparent dynamic range ($\delta$). (C) Predicted effect of residual inducer on the apparent Hill coefficient ($\eta$). (D) Predicted effect of residual inducer on the apparent half-maximal inducer concentration ($EC_{50}$).

5.3b Experimental verification of the effects of residual inducer for an inducing system in *E. coli*

To explore these predictions experimentally, we first evaluated the impact of residual inducer on a standard inducible system in the bacterium *Escherichia coli* for synthetic biology, metabolic engineering, and fundamental genetic studies: the L-arabinose-inducible P$_{BAD}$ promoter (Cox et al., 2007; Guzman et al., 1995; Mandin and Gottesman, 2009). This promoter is naturally repressed by the sensory regulator AraC through DNA-looping interactions (Lobell and Schleif, 1990). However, in the presence of L-arabinose, the AraC dimer undergoes a conformational change and recruits RNA polymerase. We employed two modified *E. coli* K-12 MG1655 strains (P$_{con}$-araE ΔaraFGH ΔaraBAD) to verify the impacts of original Hill coefficients on response parameters for different residual inducer
concentrations. The strains were transformed with the low-copy pUA66-ParaB plasmid expressing GFP under control of the P_{BAD} promoter and grown to mid-log phase in the presence of varying L-arabinose concentrations prior to flow cytometry analysis. We initially pre-incubated the cells overnight in L-arabinose as the source of residual inducer, followed by back-diluting cells into the same medium supplemented with additional amounts of L-arabinose. However, we found extremely similar responses for pre-incubated cells and the quantitative response for cells exposed to L-arabinose upon back-dilution (Supplementary Fig. S5.3), prompting us to employ the latter approach when exploring the impact of residual inducer in *E. coli*.

For the strain with η>1, we generally observed the predicted trends when varying the residual concentration of L-arabinose present in the medium (Fig. 5.2). The apparent dynamic range shrunk with increasing concentrations of residual L-arabinose (Fig. 5.2B) that began to plateau before reaching the true EC_{50} value (7.5 μM, Fig. 5.2D), in line with model predictions (Fig. 5.1B). Furthermore, the apparent response became less steep (from η = 1.62 to η = 1.03) with increasing concentrations of residual L-arabinose (Fig. 5.2C). Also paralleling model predictions, the apparent Hill coefficient approached a value of one before reaching the true EC_{50} value (Fig. 5.2C). Finally, the apparent EC_{50} were roughly constant, which was expected due to the small deviations in EC_{50} values predicted by the model at concentrations of residual L-arabinose below the true EC_{50} value (7.5 μM, Fig. 5.2D).

5.3c Assessing the effects of residual inducer for a repressing system in *E. coli*

Similar to the activating system, we used mathematical modeling to investigate the impacts of residual inducer on the response parameters. While the inducing and repressing systems exhibited similar qualitative trends (Fig.1(bottom)), we observed quantitative differences that reflect the susceptibility of each system to residual inducer. In particular, the activating system was more susceptible than the repressing system to residual inducer, as lower concentrations of residual inducer impaired the dynamic range and flattened the apparent response for η>1. Contrary to the activating system, the dynamic range and Hill coefficient of the repressing system are significantly affected by the original Hill coefficient η. While
η > 1 causes drastic fall in dynamic range with increasing residual concentrations, η ≤ 1 had less pronounced effect on the dynamic range. Overall, these results suggest that residual inducer imparts distinct quantitative effects on the perceived response curve, with expected differences between inducing and repressing systems. We next experimentally explored the impact of residual inducer on a repressing system in *E. coli*. We selected the thiamine pyrophosphate (TPP)-responsive riboswitch in the 5’ untranslated region of the *thiC* gene, which represses expression of the *thiC* operon in *E. coli* in the presence of exogenous thiamine (Winkler et al., 2002). Naturally, thiamine is enzymatically converted into TPP, which triggers the riboswitches to halt expression of the TPP biosynthetic operons (Serganov et al., 2006). We constructed a reporter plasmid (pUA66-thiC) in which GFP is translationally fused to the *thiC* riboswitch and the first 14 codons of *thiC* and is driven by a synthetic constitutive promoter (Fig. 5.2E). *E. coli* K-12 MG1655 cells transformed with the pUA66-thiC plasmid were then cultured in varying concentrations of thiamine in media already containing residual thiamine.

As shown in Figure 5.2E-H, we found close agreement between the extrapolated experimental values and the model predictions. The apparent dynamic range and Hill coefficient both decreased with increasing residual thiamine (Fig. 5.2F,G); as predicted, both parameters continued to decrease even at concentrations of residual thiamine greater than the true EC₅₀ value (4.35 μM, Fig. 5.2H). Furthermore, the apparent EC₅₀ value showed the predicted non-monotonic dependence on residual thiamine, with the trough around the true EC₅₀ value (4.35 μM, Fig. 5.2H). These findings further demonstrate the impact of residual inducer on repressing systems and the relative susceptibility of inducing and repressing systems to residual inducer.
Figure 5.2. Impact of residual inducer on titratable systems in *E. coli*. (A) Response curves for *E. coli* MG1655 P\textsubscript{con-araE} ΔaraE ΔaraBAD cells harboring pUA66-paraB with the indicated concentration of residual L-arabinose. The gray bar represents the measured autofluorescence of the *E. coli* cells. Impact of residual L-arabinose on (B) the dynamic range (δ), (C) the hill coefficient (η), and (D) the half-maximal inducer concentration (EC\textsubscript{50}) for P\textsubscript{con-araE} ΔaraE ΔaraBAD cells. (E) Response curves for *E. coli* MG1655 cells harboring pUA66-thiC with the indicated concentration of residual thiamine. Impact of residual thiamine on (F) the dynamic range (δ), (G) the Hill coefficient (η), and (H) the half-maximal inducer concentration (EC\textsubscript{50}) for MG1655 cells. Values represent the mean and S.E.M. of independent experiments with at least three separate colonies.

5.3d Assessing residual inducer for an “all-or-none” response in *E. coli*

While our modeling and experimental efforts focused on systems that yield unimodal responses, many other inducible systems are known to exhibit bistable or “all-or-none” behaviors (Novick and Weiner, 1957). This behavior is typified by the inducible system being fully or negligibly induced in single cells. One prevalent example is the P\textsubscript{BAD} promoter, which exhibits “all-or-none” behavior in response to exogenous L-arabinose particularly in strains lacking catabolic activity (Afroz et al., 2014b; Khlebnikov et al., 2001; Morgan-Kiss et al., 2002; Siegele and Hu, 1997). L-arabinose induces expression of a high-affinity (AraF/G/H) and low-affinity (AraE) transport system that import more L-arabinose into the cytoplasm, resulting in a positive feedback loop that drives full induction of the promoter.
To examine the impact of residual inducer on an “all-or-none” response, we transformed a strain of *E. coli* K-12 MG1655 unable to consume L-arabinose (Δ*araBAD*) with the pUA66-ParaB reporter plasmid. The strain was then cultured in varying concentrations of L-arabinose in medium already containing residual L-arabinose. We avoided pre-incubating the cells as part of the overnight culture to avoid induction at low L-arabinose concentrations due to extended exposure (Afroz et al., 2014b). The resulting bimodal responses were captured as dot plots to communicate the relative abundance and fluorescence of each sub-population (Figure 5.3).

At lower concentrations, the residual L-arabinose increased the sensitivity of the promoter to applied L-arabinose. This can be readily attributed to the residual L-arabinose approaching the concentration in which cells transition from the uninduced to the induced state. Once the residual L-arabinose reached the switching threshold (~3 μM), a portion of the population was fully induced even in the absence of applied L-arabinose. At higher concentrations of residual L-arabinose, the entire population was fully induced and was no longer responsive to applied L-arabinose. These findings indicate that residual inducer can sensitize an “all-or-none” system to the inducer, although excessive amounts of the residual inducer can drive the system into the fully induced state.

### 5.4 Discussion

Using mathematical and experimental approaches, we found that the presence of residual inducer can quantitatively reshape the perceived properties of an inducible expression system. The inducer generally acted to impair the dynamic range and flatten the response curve, which we observed in inducible and repressible systems in bacteria. This phenomenon is primarily mathematical in nature, as we are effectively rescaling the horizontal axis of the response curve. However, this phenomenon has physical implications when inducer present in the medium was unknown, such as tetracycline found in some animal cell culture reagents, or when the inducer cannot be removed, such as for inducers manufactured by cells. In either scenario, the presence of the residual inducer skews the perceived relationship between inducer concentrations and the output of the inducible system. This could greatly influence
how a given inducible system is viewed by the scientific community and how it is used as a research or engineering tool. While residual inducer is normally avoided when working with inducible systems, its presence can be beneficial. Specifically, the flattening of the response curve yields a more tunable response, allowing fine-tuning of expression levels with the applied inducer. The loss of the dynamic range certainly is a detriment, reflecting an inherent

Figure 5.3. Impact of residual inducer on an “all-or-none” system in E. coli. (A) Dot plots for E. coli MG1655 ΔaraBAD cells harboring pUA66-paraB with the indicated concentration of residual L-arabinose. Filled and empty circles represent the induced and uninduced subpopulations, respectively. The size of the dot reflects the proportion of cells in the subpopulation, whereas the height of the dot reflects the mean fluorescence of the subpopulation. The gray bar represents the measured autofluorescence of the E. coli cells. Values are representative of independent experiments starting with at least three separate colonies. (B) Impact of residual L-arabinose on the inducer concentration needed to induce half the population. Values represent the mean and S.E.M. of independent experiments with at least three separate colonies.
trade-off between positive (lower Hill coefficient) and negative (lower dynamic range) influences of residual inducer. Note that this trade-off marginally included the EC\textsubscript{50} value, which determines the amount of potentially costly inducer needed to control the inducible system. We did observe quantitative differences between inducing and repressing systems that impact the influence or potential utility of residual inducer. For instance, repressing systems were less sensitive than activating systems to residual inducer. This would translate into background levels of inducers exerting less of influence on the response properties—for example, when using TetON (inducing system) or TetOFF (repressing systems) with medium containing residual tetracycline. For our systems the differences could possibly arise because the activating system uses a regulatory protein whereas the repressing system acts in the posttranscriptional level. However, the modeling results with a generic equation also predicted repressing systems to be less susceptible than activating systems, supporting the generality of this finding.

We further observed that the effects of residual inducer on the response parameters depended significantly on the original Hill coefficient values. This meant that the residual inducer could have both positive or negative effects depending upon the properties of the original response. One intriguing finding is that the apparent Hill coefficient (\(\eta'\)) always converges to 1 irrespective of the original Hill coefficient value.

Overall, our work revealed a unique phenomenon associated with the presence of residual inducer. This phenomenon would not be obvious when working with any inducible system and must remain a possibility when characterizing new inducible systems or working with different growth media. However, when present, any residual inducer is expected to shape how inducible expression are perceived and employed whether in fundamental genetic studies and in synthetic biology applications.

**Acknowledgements**

We thank Caleb Wilson for constructing the thiC reporter plasmid. This work was supported by the National Institutes of Health (1R56AI103557-01A1 to C.L.B.). T.A. and C.L.B. conceived of the study; T.A. performed experiments in \textit{E. coli}; T.A. and M.L.L. performed
the mathematical modeling; T.A., and C.L.B. designed the experiments and analyzed the data; and C.L.B. and T.A. wrote the manuscript.

Conflict of Interest

The authors declare no financial or commercial conflicts of interest.
References


Chapter 5

Impact of residual inducer in titratable expression systems

(Supplementary information)

*This work will be submitted to Biotechnology Journal. Contributions to this work: Michelle Luo
## Supplementary Tables

**Table S5.1.** Strains used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Genotype</th>
<th>Source</th>
<th>Stock #</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG1655</td>
<td><em>Escherichia coli</em> K-12 substrain MG1655</td>
<td>Gottesman lab (NIH)</td>
<td>CB1</td>
</tr>
<tr>
<td>ΔaraBAD</td>
<td>MG1655 ΔaraBAD::cat</td>
<td>(Afroz et al., 2014b)</td>
<td>CB2</td>
</tr>
<tr>
<td>P_{con}araE</td>
<td>ΔaraBAD ΔaraBAD::bla</td>
<td>(Afroz et al., 2014a)</td>
<td>CB3</td>
</tr>
</tbody>
</table>

**Table S5.2.** Plasmids used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Source</th>
<th>Stock #</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUA66</td>
<td>Low-copy plasmid (pSC101 ori) with MCS upstream of <em>gfp</em></td>
<td>OpenBiosystems</td>
<td>pCB198</td>
</tr>
<tr>
<td>pUA66-ParA</td>
<td>pUA66 with the <em>araB</em> promoter</td>
<td>OpenBiosystems</td>
<td>pCB208</td>
</tr>
<tr>
<td>pUA66-thiC</td>
<td>pUA66 with synthetic promoter (BBa_J23119) and portion of the <em>thiC</em> riboswitch and the first 14 codons of the</td>
<td>This study</td>
<td>pCB404</td>
</tr>
</tbody>
</table>
**Table S5.3.** Oligonucleotides used in this study. Bases shown in red are part of the synthetic promoter (J23110). Bolded bases are the homology regions for recombination into MG1655.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>del-araBAD.fwd</td>
<td>TCCATACCCGTTTTTTTGATGGAGTGAAACGCATGGT</td>
</tr>
<tr>
<td>del-araBAD.rev</td>
<td>CCATATGAATATCCTCCTTTAG</td>
</tr>
<tr>
<td>ParaE.fwd</td>
<td>TCTGCTGTTAAATTAGGTGGTTAATAATAATCTCAAATA</td>
</tr>
<tr>
<td>J23110.rev</td>
<td>ATTCAGTAGGGAGCTGCTTT</td>
</tr>
<tr>
<td>ParaE-J23110.rev</td>
<td>GCTAGCATTTGTACCTAGGACTGAGCTAGCCGTAAAACAT</td>
</tr>
<tr>
<td>del-araFGH.fwd</td>
<td>GATAGTAACCATTTTTTTTCTGCGAGAGAGTAAGAC</td>
</tr>
<tr>
<td>del-araFGH.rev</td>
<td>GCTAGCATTTGTACCTAGGACTG</td>
</tr>
<tr>
<td>sc101.fwd</td>
<td>TTACCAATGCTTTAATCAAGT</td>
</tr>
<tr>
<td>sc101.rev</td>
<td>ATGCTTTAGCTAATACACTAAGCAT</td>
</tr>
<tr>
<td>pUA66.fwd</td>
<td>AGTAAAGGAGAGAAACTTTTCAT</td>
</tr>
<tr>
<td>pUA66.rev</td>
<td>ATTATACCTAGGACTGAGCTAGCTGCAAGTGGATGAGACG</td>
</tr>
<tr>
<td>thiC_fwd</td>
<td>CTAAGCTCTAGTCTGATAGTTATAGCTG</td>
</tr>
<tr>
<td>thiC_rev</td>
<td>GAAAAGTTCTTCTCCTTTTACTGCGGCGGTTTACGCGGCGG</td>
</tr>
</tbody>
</table>
Supplementary figures

Figure S5.1. Performance metrics based on residual inducer. (A) Parameters in the Hill equation. The plot reflects log axes. The dynamic range ($\delta$) is the ratio of maximal over minimal expression levels. The Hill coefficient ($\eta$) reflects the slope of the response curve. The half-maximal inducer concentration ($EC_{50}$) is the inducer concentration that yields the average of the maximal and minimal expression levels. (B) Illustration of the impact of residual inducer on the parameter values. Residual inducer stretches the response curve (red) to the left, influencing the value of the apparent dynamic range ($\delta'$), apparent Hill coefficient ($\eta'$), and the apparent half-maximal inducer concentration ($EC_{50}'$). Illustrative examples are shown for an inducing system (top) and a repressing system (bottom). See Figure 5.1 for more information.
Figure S5.2. Predicted impact of residual inducer for varying values of the dynamic range or the half-maximal inducer concentration. Simulations were conducted similar to those shown in Figure 5.1. Impact of varying the original dynamic range ($\delta$) on the (A) apparent dynamic range ($\delta'$), (B) apparent Hill coefficient ($\eta'$), and (C) apparent half-maximal inducer concentration ($EC_{50}'$). $\eta = 3$, $EC_{50} = 1$. Top: activating system, Bottom: repressing system. Overlapping curves are slightly offset to avoid masking the other curves. Impact of varying the original half-maximal inducer concentration ($EC_{50}$) on the (D) apparent dynamic range ($\delta'$), (E) apparent Hill coefficient ($\eta'$), and (F) apparent half-maximal inducer concentration ($EC_{50}'$). The displayed amount of residual inducer was scaled with $EC_{50}$. Top: activating system, Bottom: repressing system. Overlapping curves are slightly offset to avoid masking the other curves. $\delta = 100$, $EC_{50} = 1$. 
**Figure S5.3.** Impact of pre-incubating residual L-arabinose on the response parameters. *E. coli* $P_{\text{con-ara}} E\Delta\text{araFGH} \Delta\text{araBAD}$ cells harboring the pUA66-ParaB reporter plasmid were exposed to the indicated concentration of residual L-arabinose either in the overnight culture prior to back-dilution (red) or as part of the back-dilution (blue). See Figure 5.2 for more information. The parameter values for the apparent dynamic range ($\delta'$), apparent Hill coefficient ($\eta'$), and the apparent half-maximal inducer concentration ($EC_{50}'$) were extrapolated by fitting the experimental data to the Hill equation. Values represent the mean and S.E.M. of independent experiments starting with at least three separate colonies. The red circles are slightly offset to the right to avoid concealing the blue circles.

**Supplementary references**


Chapter 6

Future work
6.1 Single-cell responses of *E. coli* to mixtures of sugars

Over the last few decades significant amount of research has been done to find sustainable alternatives to fossil fuels. One promising avenue of research has been the engineering of microorganisms to convert plant biomass into chemicals and biofuels. However, plant biomass consists of mixtures of sugars such as L-arabinose and D-xylose and as industrially relevant microorganisms usually exhibit hierarchal utilization of sugars, the efficiency of using these microorganisms for biofuel production can be hugely affected. The first step in overcoming this bottleneck is to understand how the microorganisms respond to mixtures of sugars at the single-cell level. The single-cell analysis for eight different sugars revealed remarkably diverse responses. How will the responses change if the cells are induced by a mixture of sugars instead of a single sugar? Studies show that presence of L-arabinose, for example, represses the expression of D-xylose operon (Desai and Rao, 2010) of *E. coli* but these studies are done using bulk characterization techniques. It remains unknown how the cells can respond to sugar mixture at the single-cell level. When only D-xylose is present, cells exhibit bimodal response whereas a combination of bimodal and graded responses is observed when only L-arabinose if present. If both the sugars are present, then for some subsaturating concentrations, a complex combination of induced and uninduced cells might be observed. The induced cells could potentially be induced by both L-arabinose and D-xylose even though it is expected that fraction of cells induced by L-arabinose would be greater than those induced by D-xylose.

Also it would be intriguing to study response of *E. coli* to sugar mixtures which otherwise yield graded response when only one sugar is present. Two sugars for which *E. coli* exhibits graded response are N-acetylglucosamine and N-acetyleneuraminic acid and these sugars are among the various sugars present in mucus (Chang et al., 2004). Could cross regulation between these two sugar utilization pathways give rise to complex response instead of the graded response? Cross regulation could potentially cause the cells to give bistable or a combination of bistable and graded responses. Chang *et. al* showed that the sence of certain sugars can affect the initiation and maintenance of *E. coli* colonization in the gut. Knowledge about how the combination of sugar mixtures induce the pathways in the
single-cell level would change our perspectives about the effects of sugar concentrations on gut colonization.

6.2 Extension of single-cell studies to sugar utilization pathways of microorganisms beyond \textit{E. coli}

Nutrient utilization is essential for the growth and survival of all forms of life in this world. Similar to \textit{E. coli}, other microorganisms like thermophiles also encode inducible sugar utilization pathways to intake and catabolize various sugars. These thermophiles are often attractive industrially because, unlike \textit{E. coli}, they can degrade biomass or can grow at elevated temperatures. Since these are less characterized microorganisms, a significant amount of research is dedicated to understanding the structures and characteristics of the utilization pathways. One significant contribution to the wealth of knowledge would be the single-cell responses to various sugars. Do the diverse single-cell responses observed true for \textit{E. coli} only or are they valid for other microorganisms also? These microorganisms carry out extracellular hydrolysis of the complex carbohydrates into simple sugars and encode various transporters to carry them inside the cells. It is predicted that the presence of multiple transporters will impact the strength of positive feedback, potentially affecting the tendency of cells to exhibit bistability. Vanfossen \textit{et. al} showed that \textit{Caldicellulosiruptor saccharolyticus} can simultaneously consume a wide variety of monosaccharides even though they are not utilized at the same extent (VanFossen et al., 2009). However, it remains unknown how the cells respond to these sugars at the single-cell level. Does the non-uniform utilization arise because the sugars are utilized differently at the single-cell level? Answers to this question are crucial because it could potentially lead to elucidation of unique regulatory mechanisms. Also, it could give insights about how theses microorganisms have evolved to survive under fluctuating nutrient conditions.
6.3 Apply the rules derived for using natural sugar utilization pathways as titratable systems for non-model microorganisms

In chapter 4, we argued that natural sugar utilization pathways can potentially be used as titratable systems for non-model microorganisms. In doing so, we derived the design rules to engineer sugar utilization pathways for titratable control. However, the rules were derived after probing L-arabinose and D-xylose utilization pathways in *E. coli*. Therefore, it remains unknown whether these rules can be applied to industrially relevant, non-model microorganisms.

To investigate the applicability of the design rules to non-model microorganisms, we first need to characterize the different pathway components such as the transporters or catabolic enzymes. Alternatively, we can start with those that have the pathway components identified and characterized. We can then analyze the single-cell response of the wild-type cells to a few relevant sugars. If they exhibit bimodal, sharp response then we can apply the rules derived from our experimental and computational results to yield titratable response. We found that constitutive expression of the high affinity transporter can readily convert a bimodal response into a unimodal response. Thus we can make this one step modification to the wild type cells to get uniform response. Also, we found that the genetic alterations came with trade-off-each modification had both desirable and undesirable effects on the response parameters. It would be intriguing to find if similar observations hold for the non-model microorganisms. By co-opting an endogenous pathway as a titratable system, researchers would potentially eliminate, for example, the challenges of using inducers or regulators non-functional at the growth conditions of these non-model microorganisms.

6.4 Tuning catabolism using CRISPR-Cas system

Mathematical modeling suggested that by increasing the enzyme levels it would be possible to proceed from a bistable response to a monostable response. Deletion of the catabolic operon for the N-acetylneuraminic pathway converted the graded response into a bistable response. However, it represented the two extreme cases: -catabolism and no catabolism. Can we change levels of the catabolic enzymes such that the bifurcation region can be tuned as
predicted by the model? This could be made possible by using the Type I-E CRISPR-Cas system in *Escherichia coli*. Luo *et al.* recently showed that this endogenous CRISP-Cas system can be used for programmable transcriptional repression by deleting cas 3 gene (Luo *et al.*, 2014).

E. coli exhibits graded response to N-acetylgucosamine and deletion of catabolic operon for this pathway gives full induction because of endogenous production of the sugar. However, we can potentially use the modified CRISPR-Cas system to repress the expression of the catabolic operon. Using this system we will further confirm the modeling predictions. In addition to verifying the modeling results, this will also enable us to show how phenotypes can be selected for using tunable parameters. For the galactose network in *Saccharomyces cerevisiae*, Avendano *et. al* showed that the transition between ON and OFF states is more sensitive to the strength of the negative feedback loop (Avendaño *et al.*, 2013). The catabolic enzymes impart negative feedback to the utilization pathways. Thus we can expect that, for the N-acetylglucosamine pathway also, the transition rate between initially induced or initially uninduced state would be affected by the levels of catabolic enzymes. The change in transition rates would thereby verify how catabolism plays an important role in helping the cells to cope under fluctuating conditions.
References


