Chapter 5

Noise in Feed-forward Loops for Galactose Utilization

In this chapter we study two naturally occurring feed-forward loops that are involved in galactose metabolism and transport. Despite having network structures that are capable of a producing dynamic, temporally diverse responses we find, by measuring dynamic noise correlations, that in their natural context these feed-forward loops are inactive. By perturbing genetic conditions the activity can be restored.

5.1 Galactose Regulation

Although *E. coli* prefer glucose as a sugar source, if glucose is not present and other sugars are available, cells will turn on the machinery to metabolize these alternate sugars. The galactose network in *E. coli* contains regulatory circuitry that implements the logic `if NOT glucose AND galactose` [52]. When this logic function is true there are two classes of genes that are turned on: galactose metabolism genes and galactose transport genes, which are known collectively as the gal regulon [53]. The metabolism and transport pathways are regulated by many of the same molecular components and the network diagrams that describe which genes affect each other are nearly conserved.

Genes for galactose metabolism and transport are turned on in response to two signals: cAMP and galactose [52]. When glucose is not present, cAMP is produced in cells and binds to the global regulator, CRP (cAMP Repressor Protein). The cAMP-CRP complex functions as an activator, turning on genes in the galactose regulon. While cAMP acts as a positive signal, galactose acts as a negative signal. Galactose binds to two repressor proteins, GalR (galactose repressor) and GalS (galactose isorepressor). GalS and GalR repress transcription by modulating the α subunit of RNA polymerase when it is bound to the promoter. The addition of galactose interrupts this process, but the galactose-GalR/S complex does not necessarily dissociate from the promoter [54, 55]. Thus,

the presence of both cAMP and galactose are necessary to turn on the metabolic and transport machinery needed so that galactose can be used as a sugar source.

The three proteins CRP, GalR, and GalS control the majority of the genes responsible for the metabolism and transport of galactose. The promoter structures of genes they regulate show many similar features with minor variations from promoter to promoter (Fig. 5.1). GalS and GalR are very similar proteins: 53% of their sequence is identical and 85% is similar. They belong to a larger family of transcriptional repressors known as the GalR-LacI family. As a result of their similarity, GalR and GalS bind to many of the same DNA binding sites and have similar features. For example, both proteins are dimers [56], are autorepressed [52, 57], and are capable of repressing each other [52]. In contrast to *galR*, *galS* is activated by the cAMP-CRP complex; *galR* has a putative CRP binding site, but has repeatedly been shown to be unresponsive to CRP [52, 53]. Despite controlling many of the same targets, the binding affinities of GalS and GalR for different genes in the gal regulon are often quite specific [53, 56].

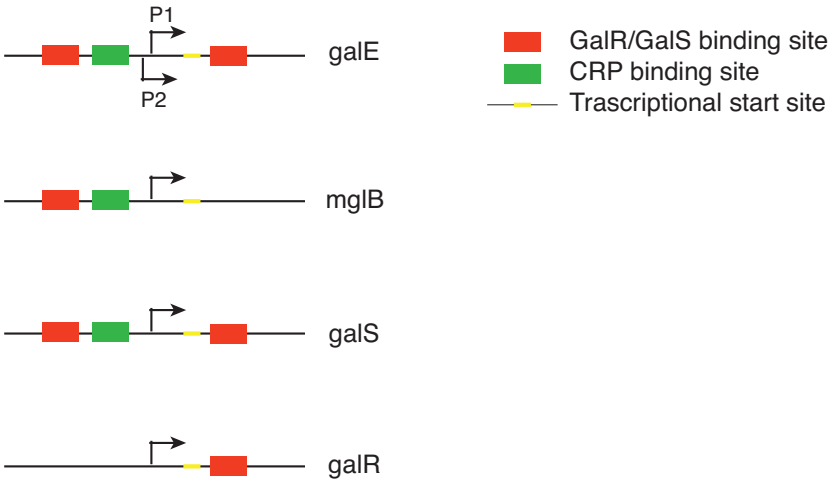

Figure 5.1: Promoter architecture for gal regulon genes. Binding sites for CRP (red boxes), GalR and GalS (green boxes), promoters (black arrows), and the transcriptional start codon (yellow bar). Many genes in the gal regulon have similar promoter and transcription factor binding sites with minor variations. This figure is based upon a diagram from [52].

Below, we go through some of the particular players in the gal regulon and describe in detail how they respond to CRP, GalR, GalS, and the signals cAMP and galactose.

5.1.1 Galactose Metabolism

There are six genes involved in the preliminary steps of galactose metabolism: *galE*, *galT*, *galK*, *galM*, *pgm*, and *galU* [53]. The first four of these genes are arranged in an operon, *galETKM*. We focus on control of this operon, referring to it as *galE* for concise notation.

galE has two promoters, P1 and P2, that control its expression (Fig. 5.2). *In vitro* studies have tested the roles of these promoters individually. In the absence of glucose, CRP activates transcription from P1 and represses it from P2. GalR and GalS play the opposite role, repressing transcription from P1 and activating it from P2 [52, 58]. In the presence of galactose, repression is relieved and transcription occurs primarily from the P1 promoter. *In vivo*, GalR plays a primary role in controlling expression of *galE*. There are two GalR binding sites O_E (external) and O_I (internal) that bracket the P1 and P2 promoters, shown in Fig. 5.3. In the absence of galactose, GalR binds to these two operators and causes the DNA to loop, obscuring the P1 and P2 promoters and inhibiting transcription. Adding galactose interrupts looping and allows for transcription from P1 and P2 as seen in the *in vitro* studies [59]. DNA looping by GalR requires formation of a structure known as the repressosome, which consists of two GalR dimers—one bound to O_E and one to O_I—and HU, a bacterial histone-like protein [60, 61]. Interestingly, operator mutation studies have shown that O_E and O_I can be replaced by LacI binding sites and full repression is maintained, suggesting that DNA looping is the major factor in repressing transcription of *galE* in the absence of galactose [58].

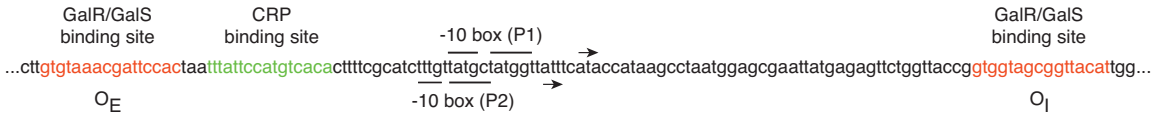

Figure 5.2: Sequence of the promoter region for *galETKM* operon. -10 boxes (and extended -10 boxes) for the two promoters are overlined (P1) and underlined (P2). Small arrows indicate transcriptional start sites for the two promoters. Red text is GalR/GalS binding sites, green is CRP.

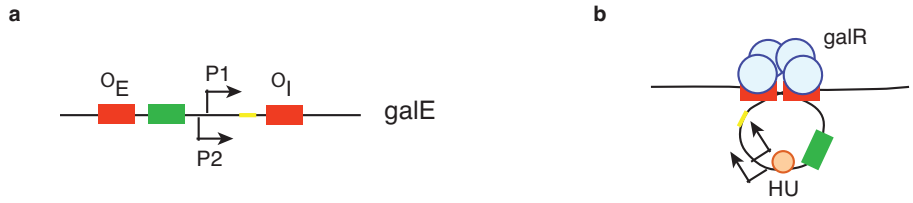

Figure 5.3: GalR repression by looping. (a) Unlooped orientation of the GalR binding operators, O_E and O_I. (b) GalR dimers tetramerize, bringing together O_E and O_I and obscuring the promoter region and transcriptional start site.

Although GalR appears to be the primary repressor of *galE*, GalS is also capable of regulating its expression. When *galR* is deleted, *galE* can be further induced by the addition of galactose, a phenomenon known as ultrainduction [62]. Deleting *galS* removes this effect. Several mechanisms have been proposed to account for this redundant regulation. Weickert and Adhya [62] postulate that GalS may serve as a backup control, alternatively, GalR and GalS may play different temporal roles, or GalR and GalS may respond differentially to levels of galactose. Work by Mangan, et

al. [63] suggests that GalS acts as part of a feed-forward loop to speed the response of galactose metabolism genes after glucose depletion.

5.1.2 Galactose Transport

Four genes are primarily responsible for transporting extracellular galactose into the cell: *mglB*, *mglA*, *mglC*, and *galP* [53]. A three-gene operon, *mglBAC*, makes up the high affinity galactose transport system, which is active when extracellular galactose is low. *mglB* is involved in binding galactose, while MglA and MglC are membrane-associated proteins. The high affinity system is primarily regulated by GalS and only weakly controlled by GalR [52]. In contrast, the low affinity galactose system is active when extracellular galactose is high and is primarily (likely soley) regulated by GalR [56]. GalP, galactose permease, is the major player in the low affinity system and is a membrane transport protein. Additional galactose transport systems exist, but are much less efficient than *mglBAC* and *galP* [53]. We focus on the high affinity transport system, *mglBAC*, abbreviated as *mglB*.

A single promoter controls expression of *mglB*. It contains a single GalR/GalS binding site and a single CRP site. The structure of the promoter is very similar to the P1 promoter on *galE* and has similar behavior: CRP activates expression in the absence of glucose, and GalR/GalS represses expression in the absence of galactose. When galactose is added, repression is relieved and *mglB* is expressed at higher levels. Unlike the *galE* promoter, because there is only a single GalR/GalS binding site DNA looping is not used to inhibit transcription. *galR* deletion experiments had little affect on *mglB* expression, while they had a strong affect on *galE*, suggesting that GalS plays a primary regulatory role [62]. This is further supported by the co-localization of the *galS* and *mglB* genes on the chromosome [53].

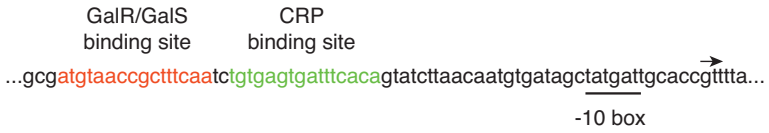

Figure 5.4: Sequence of promoter region for *mglBAC* operon. Labeling and symbols are consistent with Fig. 5.2.

5.1.3 Structure of Regulatory Networks

Control of *galE* and *mglB* is implemented by regulatory circuits with very similar structures. Both operons are the target of feed-forward loops involving CRP and GalS that respond to the signals cAMP and galactose. Fig. 5.5 contains all of the regulatory connections for controlling *galE* and *mglB* that have been proposed in the literature. Databases like RegulonDB [7] contain information

in this form, which is a useful starting point for understanding regulation and cataloging all possible interactions. However, as pointed out in [52], not all of the regulatory connections that are listed in Fig. 5.5 are necessary to produce the cellular responses that are observed *in vivo*. In particular, we show that the context in which the circuit operates is very important for determining its function.

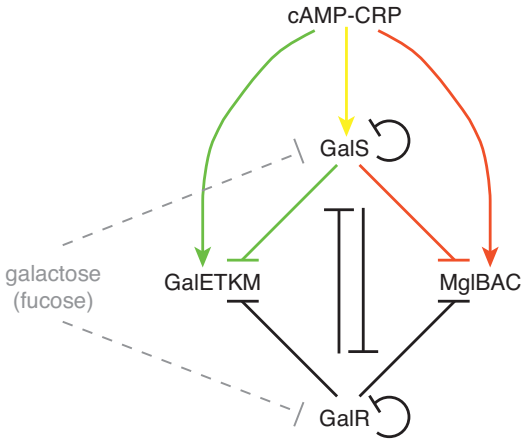

Figure 5.5: Regulatory network controlling expression of *galE* and *mglB*. Two feed-forward loops are highlighted in red and green, with a common element colored in yellow. Solid lines represent transcriptional regulation and dashed lines are non-transcriptional interactions.

5.2 Clustering of Type I Incoherent Feed-Forward Loop Responses

In Chapter 4 we saw that feed-forward loops can exhibit different types of responses depending upon the regulatory circuitry, response to signals, and system parameters. Both the feed-forward loops governing galactose metabolism and transport are Type I Incoherent feed-forward loops [12] that respond to two signals. We consider these two signals independently, though the analysis can be extended to multiple signals, as in [64]. Galactose does not affect the X-Y or X-Z connections, but inhibits repression in Y-Z, thus, in the notation from Chapter 4, the influence of galactose is type $\{0, -, 0\}$. cAMP has a positive effect on the X-Y connection and X-Z connection, but has no effect on Y-Z, so the signal's influence is $\{+, 0, +\}$.

Results from screening circuits of this type over a broad range of parameters are summarized in Fig. 5.6. The response of the feed-forward loop to a pulse in galactose is very stereotyped: for all combinations of parameters explored the circuit exhibits simple activation. A pulse in cAMP can exhibit responses ranging from simple activation to accelerated response with overshoot in reaction to an ON step in the signal. Activation without an accelerated response is the most common response, but a significant fraction (20%) of the conditions tested showed pulsing behavior.

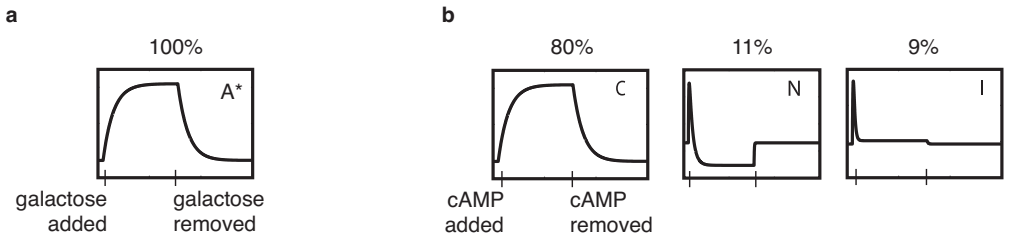

Figure 5.6: Predicted behavior of feed-forward loop target gene (*galE* or *mglB*) expression in response to pulses in (a) galactose and (b) cAMP. Numbers in title bars indicate what percentage of simulated systems fell into this cluster. Letter labels correspond to Figs. 4.7 and 4.8. A^* is the signaling inverse of A; due to symmetry not all signaling interactions were simulated, thus a signal turning ON for $\{0, -, 0\}$ is the same as a signal turning OFF for $\{0, +, 0\}$.

These results indicate that pulses in cAMP are capable of accelerating the response of *galE* and *mglB*, but galactose cannot elicit a similar temporal effect.

5.3 Context Sensitive Regulation

The literature reviewed in Sections 5.1.1–5.1.2 suggests some simplifications to the network diagram shown in Fig. 5.5. The regulatory elements controlling expression of *galE* are shown in Fig. 5.7a, where elements that exist, but do not play an active regulatory role are colored in gray. Fig. 5.7b shows the active regulatory network for *mglB*.

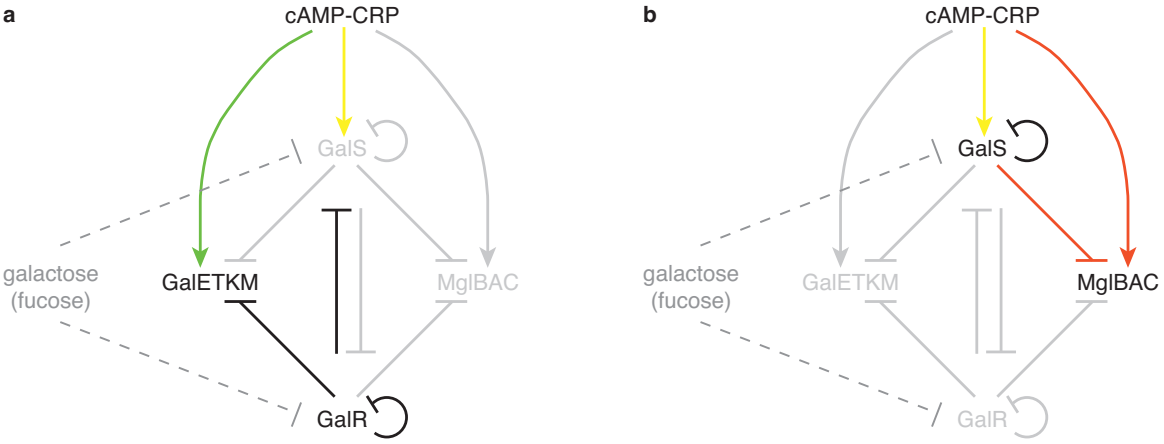

Figure 5.7: Active and inactive regulatory circuitry for control of expression in (a) *galE* and (b) *mglB* based on literature review. Inactive network components are shown in gray, all others play an active role in regulation.

Adding an inducer can change the context in which a genetic circuit is active. We ran all tests using mannose as a sugar source. This ensured that expression levels of *galS*, *galE*, and *mglB* were high enough to be visible, due to activation by cAMP-CRP [63]. Preliminary experiments indicated that mannose, as opposed to glucose, would allow for sufficient activation of several genes of interest

(Fig. 5.8). As in [63], we used fucose, a non-metabolism analog to galactose, as an inducer. Like galactose, fucose binds to GalS and GalR and inhibits ability to repress. Thus, when fucose is present autoregulation and repression are relieved, further simplifying the circuit diagrams shown in Fig. 5.7.

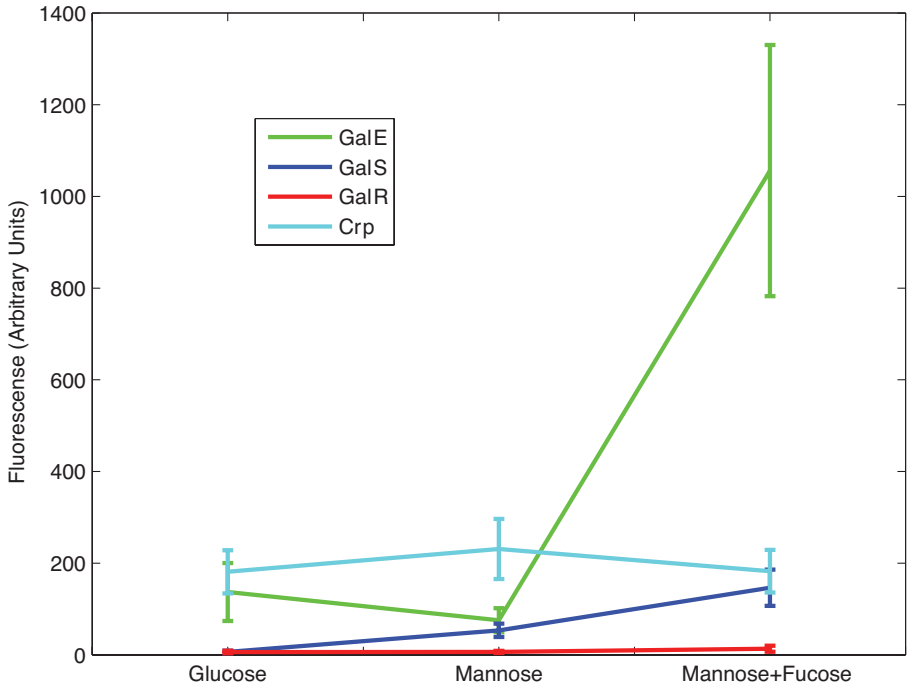

Figure 5.8: Response of P_{galS}, P_{galR}, P_{crp}, and P_{galE} to different sugars. Error bars show standard deviations. Mannose and glucose are 0.8%, fucose is 20 mM, all cells were grown in M0 (see Methods). Expression levels were measured individually using plasmid-based promoter-GFP fusions from [65]. Fluorescence data have been background subtracted. P_{mglB} is expected to have measurable expression levels based on measurements from [66].

To test the activity of the *galE* and *mglB* feed-forward loops we constructed a set of promoter-fluorescent protein fusions for pairwise measurement of gene expression from *galS*, *galE*, and *mglB*. Promoters for these genes are the same as in [65], placed upstream of the *yfp* and *cfp* genes used in the synthetic circuit described in Chapter 3. The promoters and fluorescent proteins were oriented in opposite directions (Fig. 5.9) to minimize read-through, and cloned next to a kanamycin resistance marker for selection. The synthetic constructs were integrated into the *intC* region of the MG1655 chromosome and colonies were screened for correct insertion length and then verified with sequencing.

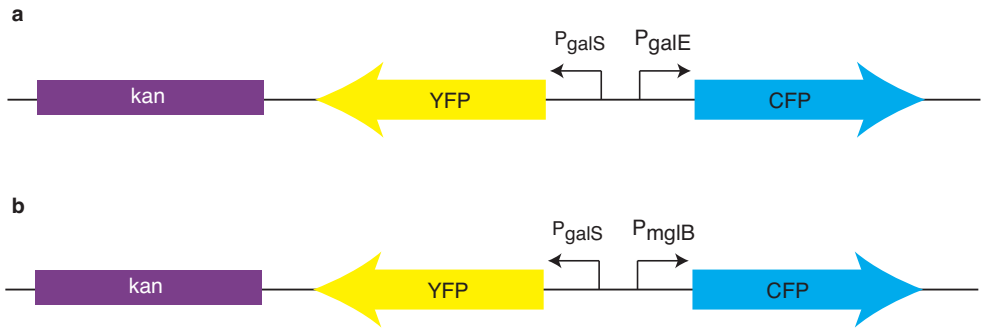

Figure 5.9: Promoter fusions for measuring (a) P_{galS} and P_{galE} and (b) P_{galS} and P_{mglB}.

5.4 Noise Correlations to Infer Activity

5.4.1 Theoretical Predictions

Promoter fusions, as compared to protein-gene fusions like that used in the synthetic circuit in Chapter 2, reduce perturbation of endogenous circuit function, and allows for signal amplification using strong ribosome-binding sites for reporter gene expression. Potential drawbacks are that intrinsic noise is no longer measured directly and if the reporter dynamics differ significantly from those of the gene of interest this will appear in the cross correlation function.

Fig. 5.10 illustrates how the expression levels of two promoters can be measured using fluorescent reporter proteins. Protein A represses B, while F is a reporter for A, and G a reporter for B.

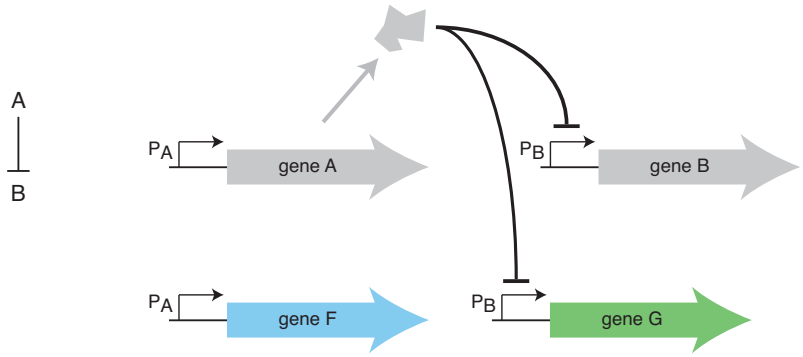

Figure 5.10: Schematic of promoter fusion for A repressing B. *gene F* and *gene G* are fluorescent proteins that can be measured to report the expression levels of promoters A and B, respectively.

This system can be modeled using the linearized approximation by

$$\begin{aligned}
\dot{a} &= -\beta a + E + I_a \\
\dot{b} &= -\beta b + g_{ab} a + E + I_b \\
\dot{f} &= -\beta_f f + E + If \\
\dot{g} &= -\beta_g g + g_{ab} a + E + I_g.
\end{aligned}$$

Here, f and g model the reporter proteins, which are expressed in the same way as the original proteins, a and b, but have different sources of intrinsic noise. The degradation rates β_f and β_g are one example of a way that the reporter dynamics could differ from the system dynamics. Fig. 5.11 shows examples of cross correlation functions generated by reporter proteins that decay more quickly than those in the original system. Cross correlation functions for the full system, assuming direct

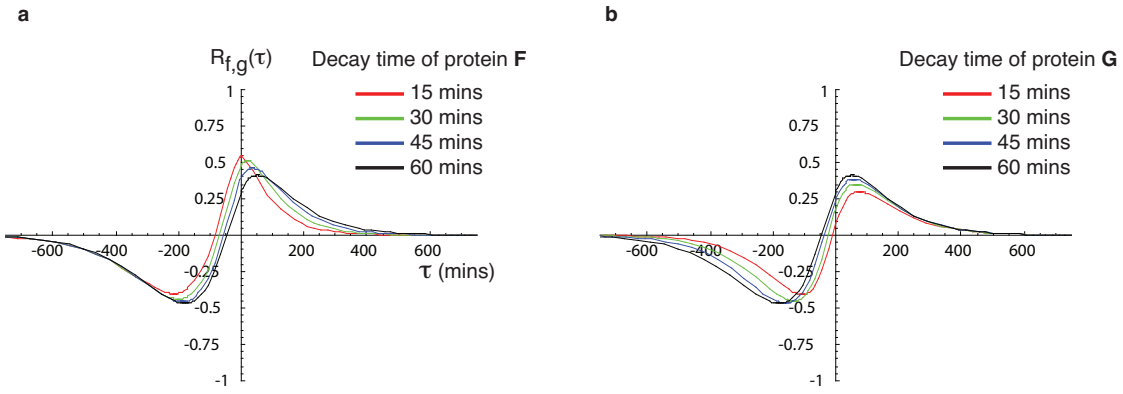

Figure 5.11: Differing reporter and system dynamics. (a) Protein F degrades more quickly than A, B, and G. (b) Protein F degrades more quickly than A, B, and G. For all calculations $g_{ab} = -0.0125$, $\theta = 0.064$, and $\beta = \beta_f = \beta_g = Log[2]/60$ unless specified in the figure caption where the decay time, T_{decay}, is used to calculate $\beta_i = Log[2]/T_{decay}$.

measurement of proteins A and B are nearly identical to those generated when $\beta = \beta_f = \beta_g$.

We used the analytic methods discussed in Chapter 2 to calculate expressions for the cross correlation function of a feed-forward loop. Fig. 5.12 shows the expected shape of the cross correlation function with and without an inducer that inhibits repression by Y. The feed-forward loop was modeled by using the linearized system of equations

$$\dot{x} = -\beta x + E + \eta_x \tag{5.1}$$

$$\dot{y} = -\beta y + g_{xy}x + E + \eta_y \tag{5.2}$$

$$\dot{z} = -\beta z + g_{xz}x + g_{yz}y + E + \eta_z, \tag{5.3}$$

with reporter proteins for Y and Z that have dynamics

$$\dot{f} = -\beta f + g_{xy}x + E + \eta_f \tag{5.4}$$

$$\dot{g} = -\beta g + g_{xz}x + g_{yz}y + E + \eta_g. \tag{5.5}$$

Parameter values are listed in the caption of Fig. 5.12. These equations represent a simplification over those in Chapter 2 because we model intrinsic noise as white noise rather than using an Ornstein-Uhlenbeck process with short correlation time. The equations in this form are simpler and give

similar results, but it is straightforward to calculate the same expressions using the more accurate intrinsic noise terms. In addition, for a complete match to the galactose feed-forward loops, the model should include autorepression by Y. This can be modeled by adding the term $g_{yy}y$ to the second equations for $\dot{y}$ and $\dot{f}$. The cross correlation between Y and Z in the active feed-forward loop

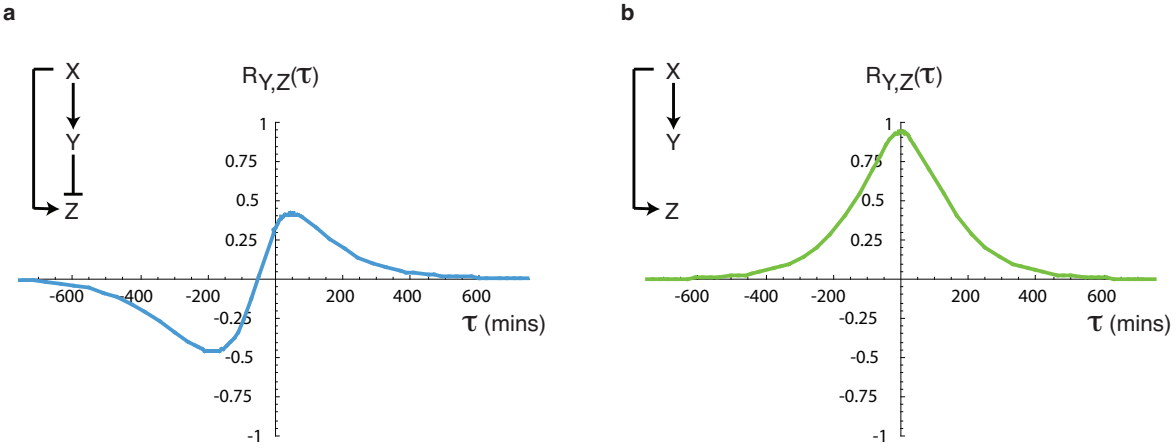

Figure 5.12: Theoretic predictions for cross correlation functions from galactose feed-forward loops (a) without and (b) with fucose. Parameter values used in these calculations are $g_{xy} = g$, $g_{xz} = g$, $g_{yz} = -g$, where $g = 0.0125$, $W_i = 1$ for $i = \{x,\ y,\ z\}$ and $W_e = 0.064$. For (b), $g_{yz} = 0$.

is similar to the response of a simple repressor, but additional positive correlation that is symmetric about zero lag is also present. The effect of X is extrinsic to both of the measured variables and consequently acts very similarly to extrinsic noise.

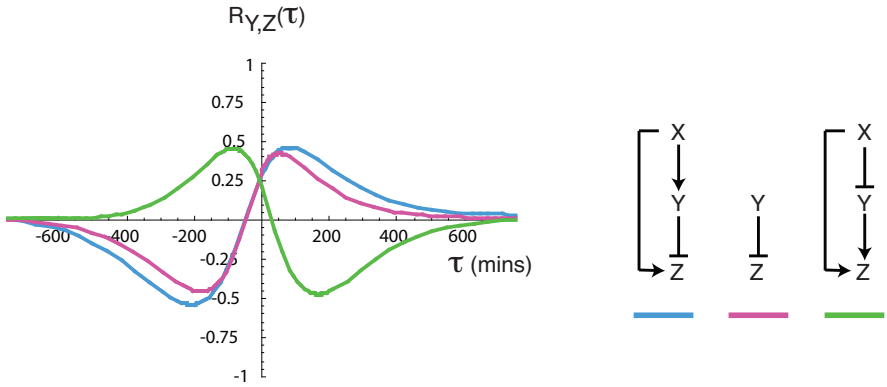

Figure 5.13: Theoretic predictions for cross correlations from alternate forms of regulation. All cross correlations are between Y and Z, regulatory architectures are indicated in the figure. The third gene circuit is an Incoherent Type 4 feed-forward loop [12]. Parameters are the same as those listed in Fig. 5.12, except $g = 0.0125$ and for the third circuit $g_{xy} = -g$ and $g_{yz} = g$.

5.4.2 Experimental Data

We measured the cross correlation between reporters for *galS* (YFP) and *galE* (CFP) in the presence and absence of fucose (Fig. 5.14). The cross correlation function without fucose has a peak at zero and is symmetric. This indicates that *galS* and *galE* are affected by some of the same noise sources and regulatory proteins, but, in these conditions, *galS* does not have a distinct regulatory effect on *galE*. Adding fucose inhibits repression by GalS and GalR, but the cross correlation curves with and without fucose are indistinguishable. These results suggest that, in the conditions we tested, GalS does not play an active regulatory role in controlling expression of *galE*.

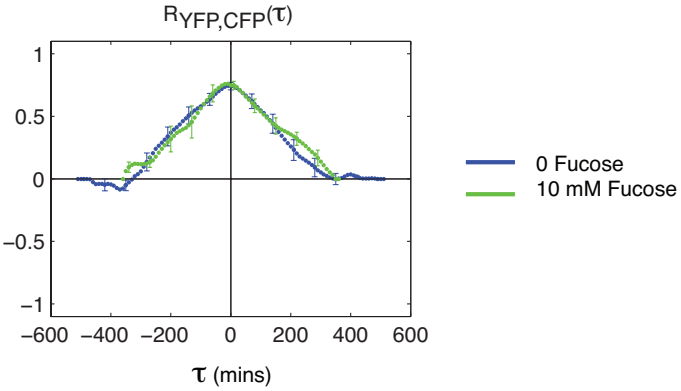

Figure 5.14: Cross correlations between P_{galS}-YFP and P_{galE}-CFP at 0 and 10 mM fucose. Error bars are standard error averaged across multiple movies. $n = 15$ movies for 0 mM fucose, $n = 8$ for 10 mM fucose.

To test whether *galE* could be controlled by *galS* in some contexts, we created a *galR* deletion strain with P_{galS} and P_{galE} reporters. The goal of this experiment was to explore the activity of the CRP/GalS/GalE feed-forward loop in the absence of repression due to looping by GalR. Without fucose the cross correlation between P_{galS}-YFP and P_{galE}-CFP is no longer symmetric and shows clear signs of repression of *galE* by GalS (Fig. 5.15). There is still a strong peak at zero lag as a result of extrinsic noise and noise in CRP, both of which are extrinsic to the measured signals. When fucose is added, repression by GalS is inhibited and the cross correlation curves are symmetric, like those seen in the presence of GalR. These data suggest that although GalS can play a regulatory role in the control of *galE* by acting a repressor of its production, in natural contexts the dominant regulatory role is played by GalR.

Note that the peak value of the cross correlation curves decrease when GalR is deleted (compare Fig. 5.14 and Fig. 5.15). This indicates that the presence of GalR was adding to the correlation between *galS* and *galE*.

Based on information in the literature, the CRP-GalS-MglB feed-forward loop should behave in a simpler fashion since it lacks looping by GalR, which was a confounding factor in the GalE feed-forward loop. Surprisingly, noise-generated cross correlations between P_{galS}-YFP and P_{mglB}-CFP

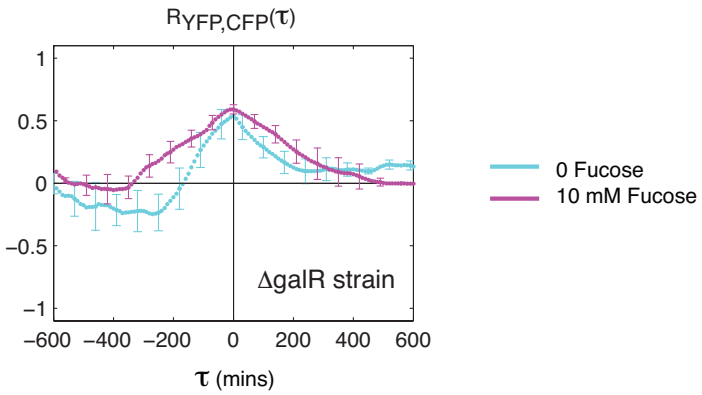

Figure 5.15: Cross correlations between P_{galS}-YFP and P_{galE}-CFP at 0 and 10 mM fucose in a *galR* deletion strain of MG1655. Error bars are standard error averaged across multiple movies. $n = 9$ movies for 0 mM fucose, $n = 9$ for 10 mM fucose.

indicate that GalS and MglB are not strongly linked, even in the absence of fucose (Fig. 5.16). Thus, it appears that in the cellular contexts that were measured for these experiments, neither galactose feed-forward loop is actively regulating its target gene. Measurements in [66] suggest that GalS may be active in other regulatory regimes.

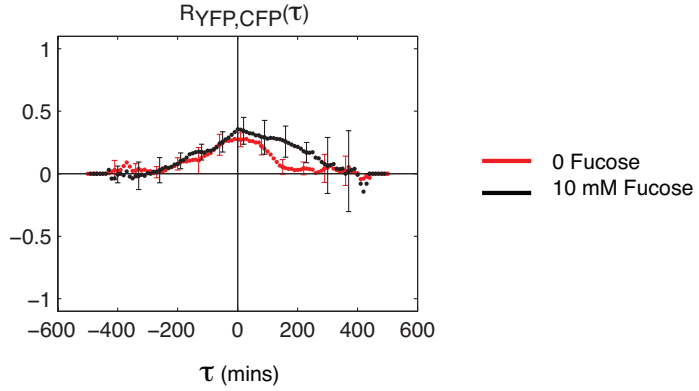

Figure 5.16: Cross correlations between P_{galS}-YFP and P_{mglB}-CFP at 0 and 10 mM fucose. Error bars are standard error averaged across multiple movies. $n = 6$ movies for 0 mM fucose, $n = 5$ for 10 mM fucose.

5.5 Methods and Characterization

5.5.1 Expression Levels from Static Data

Measurements from snapshot data are shown in Fig. 5.17. Addition of 10 mM fucose raises expression levels of all three proteins by relieving repression by GalR and GalS. When GalR is deleted, expression of GalS and GalE increase further with fucose addition, but remain low without fucose—likely

the result of residual repression by GalS.

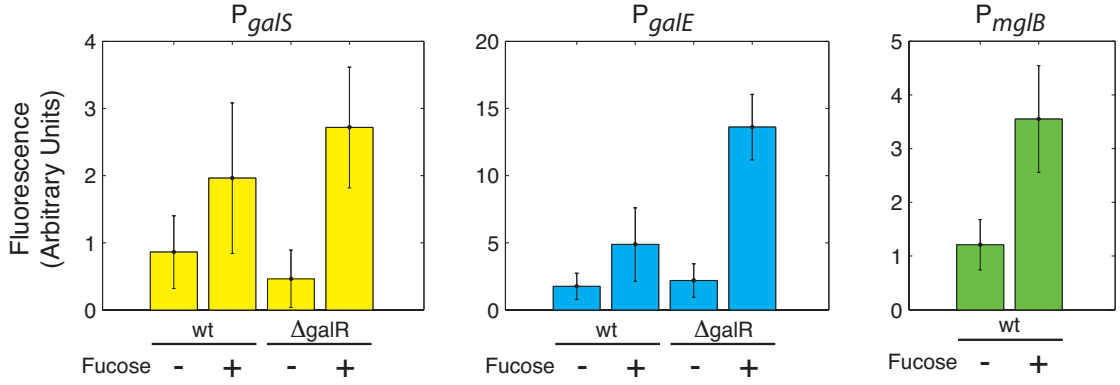

Figure 5.17: Expression levels of reporters for P_{galS}, P_{galE}, and P_{mglB} under different induction and genetic conditions. Data are background subtracted. Error bars show standard deviation over all the individual cells measured in snapshots (typically 100-200 cells). + Fucose is 10 mM, all cells are grown with mannose as a sugar source.

5.5.2 Methods

Promoter regions for *galS*, *galE*, and *mglB* were taken from plasmids in Alon Zaslaver's reporter library [65]. Promoter-fluorescent protein fusions were made with fusion PCR and verified by sequencing. The fusion PCR product was cloned into a vector with the kanamycin resistance marker and a low copy (SC101) origin of replication (pZS2 [42]). The region from kanamycin through the terminators following CFP was amplified using PCR with homology arms for *intC*

H1: $5' - \texttt{CCGTAGATTTACAGTTCGTCATGGTTCGCTTCAGATCGTTGACAGCCGCA} - 3'$

H2: $5' - \texttt{ATAGTTGTTAAGGTCGCTCACTCCACCTTCTCATCAAGCCAGTCCGCCCA} - 3'$,

and integrated into the MG1655 chromosome using recombineering [44].

galR was deleted from the MG1655 strain with chromosomally integrated P_{galS}-YFP/P_{galE}-CFP. The chloramphenicol marker from pKD3 was amplified using the PCR primers described in [67] (P1: 5' – `GTGTAGGCTGGAGCTGCTTC` – 3', P2: 5' – `ATGGGAATTAGCCATGGTCC` – 3') with homology arms for *galR* deletion from [68]:

H1: $5' - \texttt{TCCGTAACACTGAAAGAATGTAAGCGTTTACCCACTAAGGTATTTTCATG} - 3'$

H2: $5' - \texttt{TACTGGCGCTGGAATTGCTTTAACTGCGGTTAGTCGCTGGTTGCATGATG} - 3'$.

Cells were grown overnight in MO (M9 salts with 1 mM $MgS0_4$, 0.1 mM $MgCl_2$, and 30 μg/ml kanamycin) supplemented with 0.4% (w/v) glucose, 0.5% (v/v) glycerol, and 0.1% (w/v) Casamino

acids (called MON in [63]). Cultures were diluted back 1:50 in MO + 0.8% mannose (and 10 mM D-fucose, where applicable). After reaching OD 0.1–0.2, cells were further diluted and placed on a pad made of the same MO + 0.8% mannose (and 10 mM D-fucose) media as the original dilution. Cells were grown and imaged at 37°C.

Image acquisition and analysis methods were identical to those described in Chapter 3.