Transcriptional Bursting in Eukaryotic Gene Regulation:

Molecular Basis and Functional Consequences

by

Tsz-Leung To

B.S., Chemical Engineering and Biochemistry.

University of Massachusetts Amherst. 2005

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering

at the

Massachusetts Institute of Technology

September 2010

Signature of author: _____
Department of Chemical Engineering
August 6, 2010

Certified by: _____
Narendra Maheshri
Assistant Professor of Chemical Engineering
Thesis Supervisor

Accepted by: _____
William Deen
Professor of Chemical Engineering
Chairman, committee of graduate students

Transcriptional Bursting in Eukaryotic Gene Regulation: Molecular Basis and Functional Consequences

by
Tsz-Leung To

Submitted to the Department of Chemical Engineering on August 6, 2010
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Transcription of mRNA appears to occur in random, intermittent bursts in a large variety of organisms. The statistics of mRNA expression can be described by two parameters: the frequency at which bursts occur (burst frequency) and the average number of mRNA produced within each burst (burst size). The mean steady-state abundance of mRNA is the product of the burst size and burst frequency. Although the experimental evidence for bursty gene transcription is abundant, little is known about its origins and consequences. We utilize single-molecule mRNA imaging and simple stochastic kinetic models to probe and understand both the mechanistic details and functional responses of transcriptional bursting in budding yeast. At the molecular level, we show that gene-specific activators can control both burst size and burst frequency by differentially utilizing kinetically distinct promoter elements. We also recognize the importance of activator residence time and nucleosome positioning on bursting. This investigation exemplifies how we can exploit spontaneous fluctuations in gene expression to uncover the molecular mechanisms and kinetic pathways of transcriptional regulation. At the network level, we demonstrate the important phenotypic consequences of transcriptional bursting by showing how noise itself can generate a bimodal, all-or-none gene expression profile that switches spontaneously between the low and high expression states in a transcriptional positive-feedback loop. Such bimodality is a hallmark in decision-making circuitry within metabolic, developmental, and synthetic gene regulatory networks. Importantly, we prove that the bimodal responses observed in our system are not due to deterministic bistability, which is an often-stated necessary condition for all-or-none responses in positive-feedback loops. By clarifying a common misconception, this investigation provides unique biological insights into the molecular components, pathways and mechanisms controlling a measured phenotype.

Thesis Supervisor: Narendra Maheshri
Title: Assistant Professor of Chemical Engineering

Acknowledgement

First of all I deeply thank my advisor Narendra Maheshri for all his vision, teaching and mentorship. Through his unique perception of how biological research should be done, he has guided me to explore a wide range of interesting scientific problems and solid techniques. His intense yet optimistic attitude towards many aspects of life has shaped me to become the person I am. And for that, I'm grateful.

Arjun Raj deserves my special thanks for teaching me single-molecular mRNA FISH. Through our infrequent but fruitful interaction, I learned the "formal" way to do science and a stylish sense of humor.

I thank my thesis committee members: Arup Chakraborty, Alexander van Oudenaarden, and Dane Wittrup for their insights and career advice. I also thank Huayu Ding and Katie Quinn who did many of the experiments shown in Chapters 3 and 5. I would also like to acknowledge the rest of the Maheshri group: Tek Hyung Lee, CJ Zopf, Bradley Niesner, Shawn Finney-Manchester and Nick Wren for their help and inputs. I am grateful to the NSF for providing financial support through a graduate research fellowship.

I owe my deepest gratitude to my mother, sister and brother. None of this work would have been possible without their best support.

Table of Contents

1. Thesis overview

This thesis attempts to provide a dynamical view of transcriptional regulation in eukaryotes. The questions motivating this thesis are: What are the crucial parameters of the transcription process? How to change those parameters? How to link those parameters to phenotypic consequences? Yet, these apparently simple questions are difficult ones to answer. The complex and multiscale nature of biological processes makes it non-straightforward to develop quantitative models that describe the numerous interactions and components involved. This complexity is further complicated by the stochastic nature of chemical transformation inside a cell. So how can one model the transcription process to gain a good deal of scientific insight? One approach is to capture the essence of a biological process with a phenomenological model that involves just a few crucial, observable parameters. In this spirit, we utilize simple stochastic models of gene activation and transcription to probe and understand the biochemistry of eukaryotic gene expression. Experimentally describing stochastic gene expression requires the application of novel techniques for imaging gene products. We use a method for imaging individual mRNA molecules in fixed cells to obtain extremely quantitative measurements of gene expression states in budding yeast.

Although the evidence for stochastic fluctuations in gene expression is abundant, little is known about the molecular basis of these fluctuations. Most previous studies that set out to recount transcriptional dynamics concentrate on describing how gene products evolve in time and space. The molecular events and detailed kinetics receive only brief treatment. In an attempt to further develop mechanistic principles underlying stochastic

gene expression, we describe the molecular and kinetic origins of transcriptional bursting, a phenomenon widely observed in a large variety of organisms. In particular, we show that the same transcription activator can control both the frequency and magnitude of the bursts by differentially utilizing kinetically distinct promoter elements. Moreover, we uncover the importance of activator residence time and nucleosome positioning on the utility of different kinetic pathways, highlighting some unique features of protein-DNA interactions. This investigation exemplifies how we can exploit stochastic gene expression to uncover the molecular mechanisms and kinetic pathways of transcriptional regulation, and how single-cell dynamical studies can complement static biochemical methods and provide a richer description of cellular processes.

At the network level, we demonstrate the important phenotypic consequences of transcriptional bursting by showing how noise itself can generate a bimodal, all-or-none gene expression profile that switches spontaneously between the low and high expression states in a transcriptional positive-feedback loop. Such bimodality is a hallmark in decision-making circuitry within metabolic, developmental, and synthetic gene regulatory networks. Importantly, we prove that the bimodal responses observed in our system are not due to deterministic bistability, which is an often-stated necessary condition for all-or-none responses in positive feedback loops. By clarifying a common misconception, this investigation provides unique biological insights into the molecular components, pathways and mechanisms controlling a measured phenotype.

While this thesis is ultimately about fundamental aspects of transcriptional regulation, its subject is not of just academic interest but also of overwhelming practical and technological importance. The regulation of gene expression is what shapes the

modern practice of metabolic engineering, functional genomics and synthetic biology. Single cells are capable of performing a vast array of physical and chemical transformations under tight regulation at the transcriptional level. Many industries, spanning agriculture, energy, the environment, biotechnology and health care, can benefit from the controlled application of these transformations.

2. Transcriptional control overview

2.1 A molecular view of eukaryotic transcription

Transcription is arguably the most central process of life and is controlled by a complex regulatory system. It is a process in which the genetic information in DNA is processed by the synthesis of complementary mRNAs by RNA polymerases. The regulatory system of transcription triggers the production of specific gene products to meet the current functional needs of the cell.

Eukaryotes possess three different forms of RNA polymerases (I, II and III). RNA polymerase II (RNA Pol II), which is composed of 12 subunits in humans and yeast, transcribes all protein coding genes and is the main target of transcriptional regulation (Roeder *et al.*, 1969). Biochemical studies revealed the existence of multiple transcription cofactors for RNA polymerase II (Matsui *et al.*, 1980). These were termed general transcription factors (GTFs), as they are involved in the transcription of all genes. The GTFs (TFIIB, D, E, F and H) and the TATA-binding protein (TBP) help RNA Pol II recognize the start site(s) on a gene.

The promoter is a specific sequence of DNA that defines the regulatory region for RNA synthesis. The GTFs and RNA Pol II assemble at the promoter to form the pre-initiation complex (PIC) initiated by the binding of TBP to the TATA element of the promoter (Lue and Kornberg, 1987). Further studies on the promoter led to the identification of gene-specific enhancer/operator elements, which are DNA sequences that bind gene-specific activator proteins (Johnson and McKnight, 1989). However, it was found that RNA Pol II, the GTFs, and the TBP only support basal transcription and

do not respond to gene-specific activators. This observation motivated the discovery of Mediator, a complex composed of about 20 subunits (Kelleher *et al.* 1990). The role of Mediator is to relay signals from gene-specific activators to RNA Pol II and the GTFs. Other coactivators such as the TBP-associated factors (TAFs) also interact with the basal transcription machinery and regulate the binding of TBP to the TATA element.

In eukaryotes, the DNA is bound to histone proteins and packaged in higher-order forms of chromatin (Kornberg, 1974). Chromatin constitutes an additional layer of regulation between the gene-specific activators and RNA Pol II since the chromatin structure needs to be disrupted before transcription can proceed. Indeed, activators can recruit complexes that modify chromatin to allow the promoter to interact with RNA Pol II and GTFs (Figure 2.1). These complexes either have histone acetyltransferase (HAT) activity or ATP-dependent chromatin-remodeling activity (Cairns, 1998).

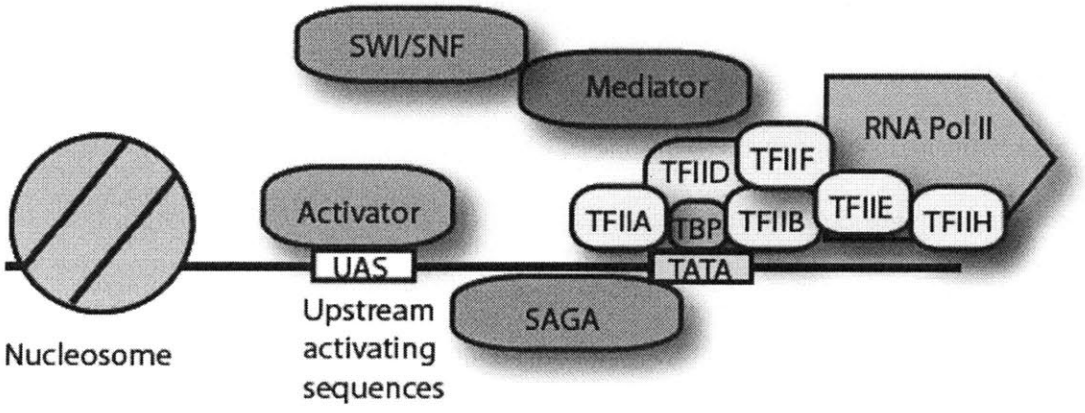


Figure 2.1. The RNA Pol II transcription machinery in yeast.

Transcription is thought to be a sequential process consisting of multiple distinct steps (Fuda *et al.*, 2009). To initiate transcription by RNA Pol II, activators must recruit the transcription machinery to the promoter and form a PIC. The ability of PIC to assemble at the promoter is also influenced by chromatin-modeling events, which govern the accessibility of activators and GTFs to their target sites. The rate of initiation is usually regarded as the rate-limiting regulatory step of transcription. After the first round of initiation, a subset of factors in the PIC remains at the promoter to form a scaffold complex that facilitates reinitiation. Reinitiation allows high rates of continuous transcription (Yudkovsky *et al.*, 2000). Transcription activity can also be regulated during elongation. The elongating RNA Pol II complex can be paused transiently and reactivated upon stimulation (Margaritis and Holstege, 2008).

Even with the knowledge of an almost complete set of proteins involved in transcription, many mechanistic aspects remain obscure. Much of the "textbook" knowledge of transcription described above comes from static biochemical and structural studies. However, these *in vitro* reconstitution studies rely on purified protein components and a controlled environment, which may not accurately reflect the complex intracellular environment. In addition, like other cellular processes, transcription is characterized as operating far out of equilibrium, and the transcription machinery maintains its organization through energy-consumptive processes. For example, interactions between protein factors and chromatin are constantly destabilized by ATP-dependent regulatory factors, such as chaperones, proteasomes and chromatin remodelers (Hager *et al.*, 2009). Indeed, recent studies show that gene-specific activators (Karpova *et al.*, 2008) and TATA-binding proteins (Sprouse *et al.*, 2008) are highly mobile *in vivo*,

in contrast to *in vitro* reconstitution studies that suggested a static and stable RNA Pol II complex (Yudkovsky *et al.*, 2000). Because transcription is dynamic and dissipative, single-cell, single molecule studies are best suited to gain new insights into the kinetics of transcription since they eliminate the population averaging that occurs in traditional *in vivo* assays. Imaging of nucleic acids and proteins *in situ* can also provide a first glimpse into the mechanisms that underlie the spatiotemporal organization of the transcription process.

2.2 A dynamical view of eukaryotic transcription

Individual cells in a genetically identical population often exhibit significant variation in gene expression. Fluctuations in gene expression can be caused by stochasticity in biochemical reactions, which becomes significant when rate-limiting reactions involve small numbers of biological molecules (Spudich and Koshland, 1976). These fluctuations are conventionally referred to as 'intrinsic noise'. Variability is also caused by differences among individual cells in cell size, stage of metabolism, abundance of regulatory molecules, etc. This type of cell-to-cell variability is referred to as 'extrinsic noise' (Elowitz *et al.*, 2002). In this thesis, we focus on intrinsic noise, which is largely determined by the underlying stochastic chemical kinetics of transcription (Raj and van Oudenaarden, 2009).

Experimental and theoretical frameworks for studying intrinsic gene expression noise have become readily available over the past decades (Maheshri and O'Shea, 2007; Raj and van Oudenaarden, 2008; Raj and van Oudenaarden, 2009). Single cell and single

molecule studies have revealed that gene activation occurs in random intermittent transcriptional bursts. In particular, Raj *et al.* (2006) were able to directly visualize mRNA in HeLa cells using fluorescent DNA probes and suggested that fluctuations in gene activation was intrinsically random and was largely unaffected by extrinsic factors. The mRNA molecules were produced in transcriptional bursts as if the gene was randomly toggling between transcriptionally active and inactive states. The statistics of mRNA expression can be described by two parameters: the frequency at which bursts occur (burst frequency) and the average number of mRNA produced within each burst (burst size). The mean steady-state abundance of mRNA is the product of the burst size and burst frequency. It should be noted that bursting does not always occur – mRNA statistics from a few housekeeping genes in yeast suggests those genes spend significant time in the active state where they produce mRNA in a Poissonian manner (Zenklusen *et al.*, 2008).

A possible origin of transcriptional bursting is chromatin remodeling: when the surrounding chromatin is in an open, acetylated state, the gene can transcribe effectively, whereas when chromatin is condensed, transcription is also repressed. This seems to be the case in higher eukaryotes as Raj *et al.* (2006) found that bursts are correlated between proximal genes, and suggested long-range chromatin remodeling as the reason for the correlations. However, there is still no direct evidence that chromatin remodeling is responsible for stochastic changes in gene activity. Although there are correlations between noisy expression of a particular promoter and the lack of chromatin-remodeling agents (Raser and O'Shea, 2004), global studies of noise in yeast have concluded that chromatin remodeling/modifying complexes are neither necessary nor sufficient for gene

expression to be noisy (Newman *et al.*, 2006; Bar-Even *el al.*, 2006). In fact, local factors such as number and placement of binding sites and TATA elements can also control noise (Blake *et al.*, 2006; Murphy *et al.*, 2007). Therefore, another possible origin of bursting is the variability in the recruitment and assembly of the RNA Pol II transcription complex.

Bursting is conveniently attributed to infrequent promoter fluctuations between an inactive state and active state (Raj *et al.*, 2006). However, this is mechanistically an oversimplification as the promoter likely adopts a series of different states involving binding of various gene-specific and general transcriptional machinery that lead to productive transcription. What does it mean to be an "active" promoter state? One idea is to consider the active promoter state as one where rapid initiation and reinitiation is possible. For example, for most inducible RNA Pol II-dependent promoters, transcriptional initiation is often rate-limiting and hence the active promoter state corresponds to the successful formation of a PIC (Hahn, 1998). In this case, the burst frequency corresponds to the initiation frequency. The burst size is related to the stability of the reinitiation intermediate (Yudkovsky *et al.*, 2000) formed at the promoter, and the rate of reinitiation (Figure 2.2).

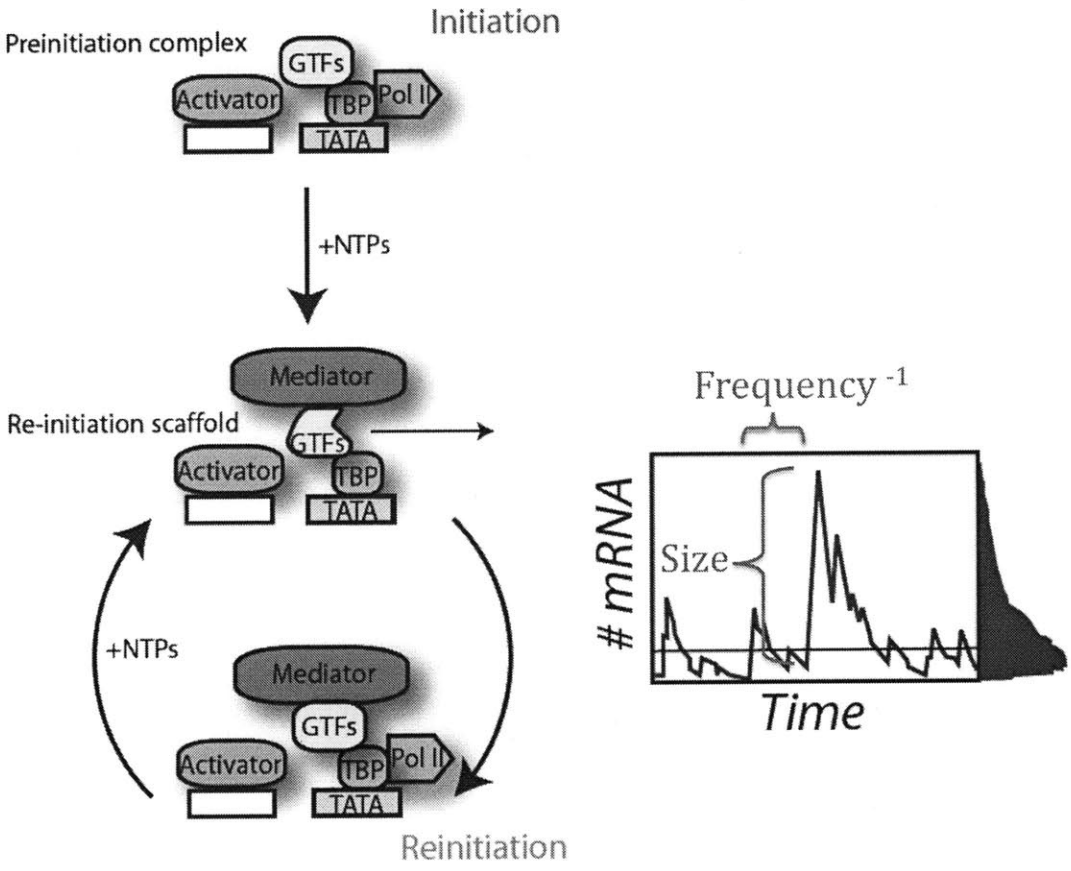


Figure 2.2. The initiation-reinitiation model underlying bursty gene transcription.

2.3 References

Bar-Even A, *et al* (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* 38: 636-643.

Blake WJ, *et al* (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* 24: 853-865.

Cairns BR (1998) Chromatin remodeling machines: Similar motors, ulterior motives. *Trends Biochem Sci* 23: 20-25.

Elowitz MB, Levine AJ, Siggia ED & Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297: 1183-1186.

Fuda NJ, Ardehali MB & Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461: 186-192.

Hager GL, McNally JG & Misteli T (2009) Transcription dynamics. *Mol Cell* 35: 741-753.

Hahn S (1998) Activation and the role of reinitiation in the control of transcription by RNA polymerase II. *Cold Spring Harb Symp Quant Biol* 63: 181-188.

Johnson PF & McKnight SL (1989) Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* 58: 799-839.

Karpova TS, *et al* (2008) Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science* 319: 466-469.

Kelleher RJ,3rd, Flanagan PM & Kornberg RD (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 61: 1209-1215.

Kornberg RD (1974) Chromatin structure: A repeating unit of histones and DNA. *Science* 184: 868-871.

Lue NF & Kornberg RD (1987) Accurate initiation at RNA polymerase II promoters in extracts from saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 84: 8839-8843.

Maheshri N & O'Shea EK (2007) Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* 36: 413-434.

Margaritis T & Holstege FC (2008) Poised RNA polymerase II gives pause for thought. *Cell* 133: 581-584.

Matsui T, Segall J, Weil PA & Roeder RG (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem* 255: 11992-11996.

Murphy KF, Balazsi G & Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A* 104: 12726-12731.

Newman JR, *et al* (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* 441: 840-846.

Raj A, Peskin CS, Tranchina D, Vargas DY & Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4:

Raj A & van Oudenaarden A (2009) Single-molecule approaches to stochastic gene expression. *Annu Rev Biophys* 38: 255-270.

Raj A & van Oudenaarden A (2008) Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135: 216-226.

Raser JM & O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304: 1811-1814.

Roeder RG & Rutter WJ (1969) Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224: 234-237.

Sprouse RO, *et al* (2008) Regulation of TATA-binding protein dynamics in living yeast cells. *Proc Natl Acad Sci U S A* 105: 13304-13308.

Spudich JL & Koshland DE,Jr (1976) Non-genetic individuality: Chance in the single cell. *Nature* 262: 467-471.

Yudkovsky N, Ranish JA & Hahn S (2000) A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408: 225-229.

Zenklusen D, Larson DR & Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* 15: 1263-1271.

3. Kinetic roles of gene-specific activators and promoter architecture revealed by bursty gene transcription

3.1 Introduction

Biological processes, such as transcription, follow the laws of chemistry. Defining the reaction steps and kinetic pathways involved in transcription programs remains an unmet challenge (Fuda *et al.*, 2009). A question that is central to the understanding of transcriptional biochemistry is: how does a gene-specific activator exert its positive regulatory effect on the selected promoter? Ultimately, environmental signals modulate transcriptional regulation through gene-specific activators. Activators are now viewed as rheostats or actuators rather than switches. With a large variety of coactivator protein complexes, activators often control transcription in a more elaborate, precise manner (Figure 3.1). Indeed, previous biochemical studies show that activators can regulate transcription through modification of chromatin template and/or direct influences on PIC formation and function (Roeder, 2004). However, enhancers and operators are often located at distant sites and it remains unclear how bound activators modulate the rates of PIC formation, initiation, and reinitiation. In particular, we do not fully understand how activators control multiple kinetic steps in the transcription process, and how promoter architecture and properties influence the kinetic roles of activators.

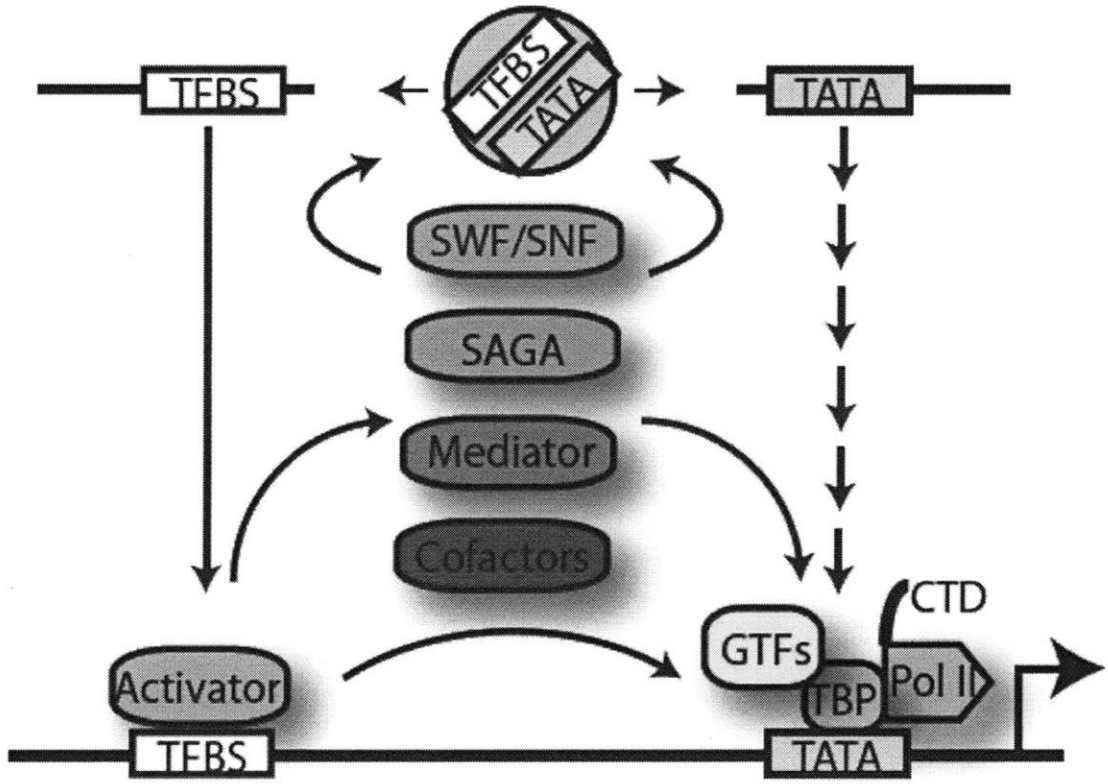


Figure 3.1. Roles of activators. Activators can interact with various coactivators to facilitate chromatin remodeling and modulate the formation and function of the general transcription machinery.

How transcription activators control the stochastic kinetics of intrinsic biochemical reactions determines the resulting expression heterogeneity. For example, activators could increase population-averaged expression by either increasing the transition from inactive to active states or stabilizing the active state. The resulting heterogeneity in expression is dictated by the specific kinetic role of the activator. The effect of activators on both the mean level of expression and expression heterogeneity is succinctly and quantitatively described by their effect on the kinetic parameters in the two-state stochastic model detailed in Section 3.2.3. A bursty eukaryotic promoter rarely

makes transitions to the active state (at a rate denoted as burst frequency), transcribing a number of mRNAs (burst size) over the short period of active state. If we were to coarse-grain the events involved in transcription, the burst frequency corresponds to the initiation frequency and the burst size is related to the rate and duration of reinitiation. In yeast, gene-specific activators function mostly by recruiting coactivator complexes to the promoter region (Ptashne and Gann, 1997). Coactivator complexes have been found to regulate both initiation and reinitiation (Fuda *et al.*, 2009). For example, the SAGA complex is responsible for chromatin modification, elongation, and mRNA export; whereas the Mediator complex can interact with and stabilize the GTF's and phosphorylate the C-terminal domain (CTD) of RNA Pol II. Acidic activators such as tTA can recruit both the SAGA and Mediator complexes (Weake and Workman, 2010), and thereby increase gene expression by regulating burst frequency, burst size, or both. Promoter binding and recruitment of various coactivators could occur at different thresholds of activator levels, resulting in complicated changes in mRNA statistics.

Inferring kinetics from steady-state mRNA statistics offers a tremendous opportunity to determine how activators and various *cis*-factors influence initiation and reinitiation. Many transcriptional activators in yeast appear to function by modulating the burst frequency by increasing the rate of transcriptional initiation (Bar-Even *et al.*, 2006; Newman *et al.*, 2006). However, activators can also modulate the burst size, as shown in a mammalian study (Raj *et al.*, 2006). In this work, we reconcile these past findings by showing experimentally that activators can affect both burst size and burst frequency, depending on the TATA configuration and the activator binding site affinity

within a set of synthetic promoters. In particular, we show that in this case burst size regulation requires both a strong activator binding site and a canonical TATA element.

3.2 Materials and methods

3.2.1 Experimental system: the Tet-OFF system

We employ the synthetic inducible Tet-OFF system, adapted for budding yeast (Gari Yeast 1997). TetO promoters are fusions of two palindromic tet operator (tetO) binding sites to a core *CYC1* TATA-containing promoter. The tet-transactivator (tTA), a fusion of the prokaryotic tet repressor and a *Herpes simplex* VP16 activation domain (Gossen and Bujard 1992), can bind tetO and activate expression. Doxycycline-bound tTA cannot bind tetO, allowing tuning of functional tTA levels (Figure 3.2). Synthetic genetic components can be treated as physical models or even replicas of existing natural systems. One advantage of choosing synthetic components over natural ones is that combinatorial libraries of many possible variants can be investigated simultaneously without impairing cellular functions.

3.2.2 Single-molecule mRNA FISH

Time-lapse microscopy provides an accurate determination of the kinetics of gene expression but it is experimentally challenging and low throughput. Instead, we can conveniently infer the kinetics of transcription from static snapshots obtained at steady-state using a stochastic model of gene expression (see Section 3.2.3). Although fluorescent proteins enable us to monitor gene expression at the single-cell level, they

have low sensitivity, and are only detectable when hundreds and thousands of proteins are present in a single cell. In addition, the fluorescence signal emitted by a single cell is of arbitrary unit. A reference standard is required to convert the fluorescence signal into the number of fluorescent proteins in a cell. Furthermore, the slow decay rate of fluorescent proteins can effectively time-average temporal fluctuations due to rapid changes in gene activation and inactivation and obscure the underlying dynamics. We can circumvent these setbacks by directly detecting single mRNA molecules in a single cell. Indeed, methods for obtaining single-cell, single-molecule mRNA statistics are readily available (Raj and van Oudenaarden, 2009). In this work, we employ the fluorescence *in situ* hybridization (FISH) method for detecting single mRNA molecules in fixed yeast cells (Raj *et al.*, 2008). Briefly, we use a large set of fluorescently-labeled 20-mer oligonucleotide probes that bind to the mRNA of the gene of interest. The specific binding of multiple probes to a single mRNA molecule results in an intensely bright, diffraction-limited spot detectable with a conventional light microscope (Figure 3.2). Importantly, this method also captures the spatial information about the location of the mRNAs detected.

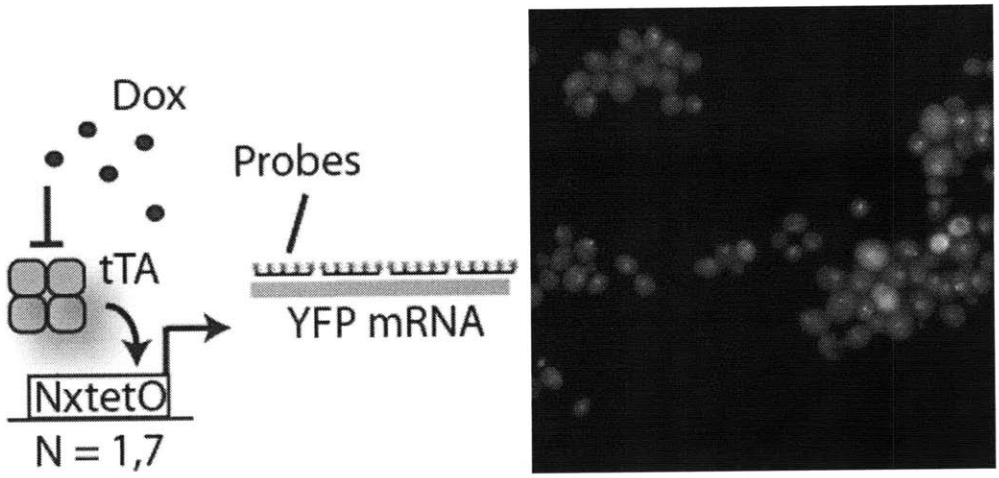


Figure 3.2. Bursty gene transcription in the Tet-OFF system. The transcriptional dynamics of the Tet-OFF system can be probed with single molecule mRNA FISH.

3.2.3 Stochastic models of gene regulation

As discussed in Section 3.2.2, single-molecule mRNA detections *in vivo* provide us a tremendous opportunity to study the detailed dynamics of the transcription process. To interpret steady-state, single-cell mRNA distributions, we need to model the stochastic chemical kinetics of the transcription process. The burst statistics of mRNA can determined based on the reaction scheme of (Raj *et al.*, 2006)

$$\begin{aligned} \mathrm{I} &\xrightarrow{\lambda} \mathrm{A} \\ \mathrm{A} &\xrightarrow{\gamma} \mathrm{I} \\ \mathrm{A} &\xrightarrow{\mu} \mathrm{A} + \mathrm{M} \\ \mathrm{M} &\xrightarrow{\delta} \emptyset \end{aligned} \qquad [3.1]$$

where I denotes an inactive promoter state, A denotes an active promoter state, and M denotes a mRNA. In the reaction scheme given by [3.1], the promoter fluctuates between

an inactive and active state. In the active state, it is capable of generating mRNA M that is eventually degraded. Raj *et al.* (2006) found a steady-state solution to the Chemical Master Equation describing this set of reactions. However, they also focused on a regime where μ is larger than the other rates in the system. Here, mRNA number is reasonably large and we consider mRNA production and degradation as deterministic, and all the randomness comes from promoter fluctuations. The steady-state solution is then given by the Beta distribution:

[3.2] $$p(\hat{x}) = \frac{\Gamma(\tilde{\lambda}+\tilde{\gamma})}{\Gamma(\tilde{\gamma})\Gamma(\tilde{\lambda})}\frac{1}{\tilde{\mu}}\hat{x}^{\tilde{\lambda}-1}(1-\hat{x})^{\tilde{\gamma}-1}$$

All rates ($\tilde{\lambda} = \lambda/\delta, \tilde{\gamma} = \gamma/\delta, \tilde{\mu} = \mu/\delta$) have been non-dimensionalized with respect to the mRNA degradation rate, δ, and are denoted with a tilde. Here, $\tilde{\lambda}$ is the burst frequency, $\tilde{\mu}/\tilde{\gamma}$ is the burst size. The continuous variable $\hat{x} = x/\mu$ represents a non-dimensional mRNA quantity, normalized by the maximum amount of mRNA possible. If gene activation is rare compared to inactivation, and promoter fluctuations are faster than the mRNA lifetime, the mRNA distribution can be described by a two-parameter Gamma distribution:

[3.3] $$P(x) = \frac{x^{\tilde{\lambda}-1}e^{-x/\tilde{\mu}/\tilde{\gamma}}}{(\tilde{\mu}/\tilde{\gamma})^{\tilde{\lambda}}\Gamma(\tilde{\lambda})}$$

This is equivalent to the result of (Friedman *et al.*, 2006) in which the same argument has also been applied to translational bursts where rare mRNA production events result in bursts of protein expression. We again emphasize that in this model, mRNA fluctuations are considered negligible and all the noise comes from transcriptional bursts due to

promoter fluctuations. Note when $\tilde{\lambda}, \tilde{\gamma}$ are both < 1, this distribution is bimodal – i.e. when promoter fluctuations are slower than the mRNA degradation rate, the mRNA distribution, even in the absence of feedback, will be bimodal. Finally, the two-parameter Gamma distribution is equivalent to the discrete negative binomial distribution (Shahrezaei and Swain, 2008):

[3.4] $$p(n) = \frac{\Gamma(\tilde{\lambda} + n)}{\Gamma(n+1)\Gamma(\tilde{\lambda})}\left(\frac{\tilde{\mu}/\tilde{\gamma}}{1+\tilde{\mu}/\tilde{\gamma}}\right)^{n}\left(1-\frac{\tilde{\mu}/\tilde{\gamma}}{1+\tilde{\mu}/\tilde{\gamma}}\right)^{\tilde{\lambda}}$$

3.2.4 Parameter estimation and uncertainty analysis for stochastic models

To illustrate how we can combine mRNA FISH measurements with a stochastic model of gene expression to determine the burst statistics of mRNA, we obtained the steady-state mRNA statistics of two 1xtetO promoter constructs under basal expression. The two constructs were integrated at the *leu2* and *ura*3 loci, respectively. FISH measurements were done in 6 replicate samples per strain. The resulting probability mass functions (PMFs) are plotted in Figure 3.3.

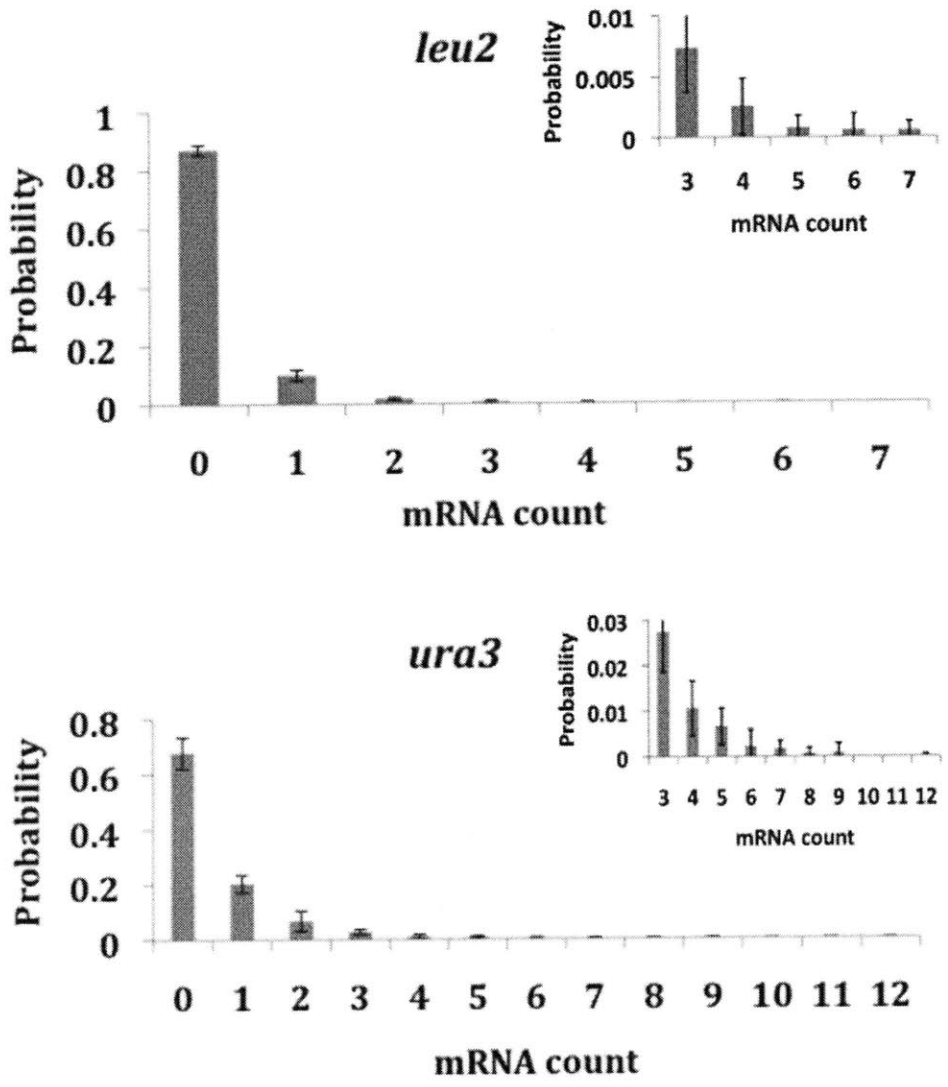


Figure 3.3. Steady-state mRNA distributions of the 1xtetO promoter under basal expression. The wild-type 1xtetO promoter was integrated at the *leu2* and *ura3* loci, respectively. Shown are the probability mass functions obtained by mRNA FISH. Error bars are standard deviations of 6 replicate samples.

Although simpler analytical expressions for the mRNA distribution are available, we first fit the FISH data sets to numerical simulations of this reaction scheme in [3.1], using maximum likelihood estimation (MLE). To do so, we first computed the mRNA distributions based on the scheme using a finite Markov approach under no assumptions

(Munsky *et al.*, 2006). Then, we used maximum likelihood to estimate the three parameters $\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}$. As in the mammalian study (Raj *et al.*, 2006), close examination of the fits showed a ridge of constant likelihood for constant $\tilde{\mu}/\tilde{\gamma}$, implying that a two-parameter negative binomial model was appropriate and activation was rare compared to inactivation (see Figure 4.4 for an example). It should be noted that although there is good agreement between parameter estimates from the full and negative binomial models, the negative binomial model seems to predict higher burst frequencies but lower burst sizes in general (but not always). Such systematic discrepancy will be the subject of future research.

Assuming a negative binomial model, the burst size and frequency can be extracted simply from the moments of this distribution:

[3.5] $$\tilde{\lambda} = \frac{\mu^2}{\sigma^2 - \mu}$$

[3.6] $$\tilde{\mu}/\tilde{\gamma} = \frac{\sigma^2}{\mu} - 1$$

Using this method, we found that the *ura3* construct has a burst frequency of 0.55 and a burst size of 0.97, whereas the *leu2* construct has a burst frequency of 0.24 and a burst size of 0.75. In general, the method of matching moments and MLE yield similar parameter estimates for the negative binomial model. An advantage of estimating parameters from moments is that we can propagate the uncertainty in the mRNA measurements into the uncertainty in parameter estimates. To quantify the technical and biological variations, we propagated the error in mRNA measurements across the 6

replicate samples to the first two moments of the PMF, and then propagated the error in the moments to the burst size and burst frequency. For the two promoter constructs under investigation, the error in burst frequency and burst size estimates due to technical and biological variations is around 30%.

The method of matching moments returns a uniquely defined burst frequency and burst size for a given sample. When MLE was used, however, we observed a ridge of similar likelihood scores for a constant multiple of burst frequency and burst size. This is not surprising since the estimates are constrained by the mean expression, which is equal to burst frequency times burst size. As such, there is often anti-correlation between the burst frequency and burst size estimates, which leads to a loss of parameter determinability. Nevertheless, by closely examining the likelihood scores we found that the uncertainty in the parameter estimates due to fitting error is substantially smaller than the 30% error due to technical and biological variations.

To assess sampling errors, we performed bootstrapping to determine the variations in burst frequency and burst size due to random sampling. The bootstrap statistics suggest that the errors in burst frequency and burst size are less than 20% for most samples. Hence, the dominant source of error is technical and biological in nature, which is estimated to be 30%. Therefore, the mRNA FISH results in this work will be presented with this 30% error.

3.2.5 Position specific affinity matrix (PSAM) for tetO mutants

We designed the tetO mutants in this study using the information from an *in vivo* single base-pair mutagenesis study (Sizemore *et al.*, 1990). We estimated the changes in Gibbs' free energy due to single point mutations, and constructed the position specific

affinity matrix (PSAM) for the relative free energy changes at different positions (Figure 3.4). The distribution of dissociation constant, K_D, of the mutant can be predicted using the PSAM and an additive model which assumes that base-pairs at different positions contribute independently to the free energy of protein-DNA interaction:

[3.7]
$$\frac{K_D(S_{mut})}{K_D(S_{ref})} = \left(\prod_{j=1}^{L} \exp\left(\Delta\Delta G_{jb} / kT \right) \right)^{-1}$$

where $\Delta\Delta G_{jb}$ is the energy change due to point mutation to base b at position j, L is the length of the sequence, $K_D(S_{ref})$ is the dissociation constant of the wild-type sequence , and $K_D(S_{mut})$ is the dissociation constant of the mutant.

	Position					
	1	2	3	4	5	6
A	-0.32	-0.48	0	-0.20	0	-0.02
T	0	-0.23	-1.15	0	-0.20	-0.41
C	-0.32	-2.32	-0.87	-0.46	-2.00	-0.44
G	-0.14	0	-1.66	-1.36	-2.00	0

Figure 3.4. The position specific affinity matrix (PSAM) for the tetO operator. The sequence of the wild-type tetO operator is shown (top). The relative free energy changes (in kT units) at different positions are based on Sizemore *et al.* (1990).

Additional materials and methods can be found in the Appendices.

3.3 Results

3.3.1 Activators control burst size or burst frequency depending on TATA elements

We titrated tTA activators using doxycycline and measured the quantitative response of various tetO promoter variants at the single cell level in hundreds of cells using the mRNA FISH assay (Section 3.2.2). For each condition and strain, we grew cells for at least 6 doublings to ensure a steady-state had been reached. In principle, to accurately measure the fast kinetics associated with promoter fluctuations, the intrinsic noise should be decoupled from extrinsic noise and analyzed to determine the burst frequency and burst size. Although extrinsic noise can be significant at the protein levels, mRNA noise in eukaryotes appears to be primarily intrinsic in nature (Raj *et al.*, 2006, Maamar *et al.*, 2007). Therefore, we expect the contribution of extrinsic factors to mRNA noise to be negligible.

To obtain the kinetic parameters we fit the measured mRNA distributions arising from each titration to a negative binomial distribution, the solution to the simplified bursting model (Section 3.2.3). This model is only strictly applicable when the inactivation rate is larger than the mRNA degradation rate and much larger than the activation rate. For each titration point, we inferred the burst frequency and burst size using the negative binomial model and the method of matching moments (Section 3.2.4). To summarize the effect of activator levels on the two kinetic parameters, we plot the burst size versus the burst frequency for individual samples in the titration series (Figure 3.5).

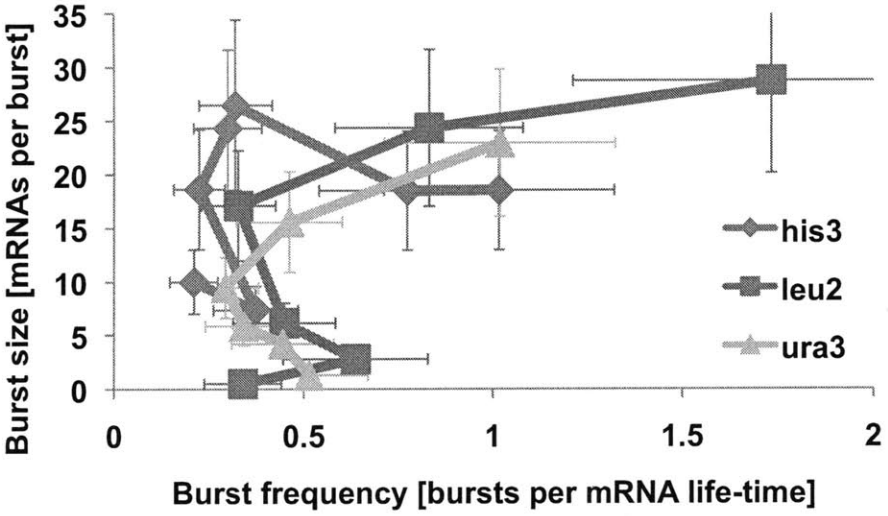


Figure 3.5. Activators increase gene expression level by a combined burst frequency and size regulation. The wild-type 1xtetO promoter was integrated at three different loci (*his3*, *leu2*, *ura3*). An *ADH1* promoter was used to constitutively express tTA. The active tTA level was modulated by 100-3000 ng/mL doxycycline. The burst frequency and burst size were inferred by fitting the steady-state mRNA distributions to the 2-state stochastic model (Section 3.2.3). Error bars represent the estimated technical and biological errors (Section 3.2.4).

We began by determining the burst statistics of mRNA for a wild-type 1xtetO integrated at the *his3* locus. At low levels, tTA activators exclusively regulate burst frequency (initiation). At intermediate levels, activators regulate mainly burst size (reinitiation). Importantly, in this regime, the burst frequency remains low and almost constant. At high levels, activators increase primarily the burst frequency by regulating initiation. The burst size in this regime is high but constant. The same qualitative trends are also observed when the same promoter is integrated at the *ura3* or *leu2* loci (Figure 3.5). A previous study of the same promoter (To and Maheshri, 2010) found that

activators primarily affected burst frequency but measurements were made at higher levels of activator. It is only by titrating activators at lower levels and using promoter variants that we were able to clearly discern the two regulatory modes.

Interestingly, we never observed a simultaneous increase in burst size and burst frequency when increasing expression from low to intermediate levels in the 1xtetO promoter. This observation points to the possibility that burst frequency and burst size are independently regulated via different pathways. Evidence for multiple initiation pathways exists. For example, the yeast *HIS3* gene contains two distinct TATA elements: a canonical TATA element for regulated expression (T_R) and a weaker non-canonical element for constitutive expression (T_C). The two TATA elements correspond to different transcription start sites and are differentially utilized at different levels of *HIS3* expression (Struhl, 1986; Iyer and Struhl, 1995). Indeed, the yeast *CYC1* core promoter also contains several functional TATA elements (Hahn *et al.*, 1985) although it shows a strong preference for utilizing the upstream canonical TATA elements (Li and Sherman, 1991). Since both canonical and non-canonical TATA elements appear to be able to recruit TBP and TFIID (Hahn *et al.*, 1989), the origins and functional consequences of this redundancy are not clear. Furthermore, the kinetic properties associated with the distinct initiation pathways have not been reported.

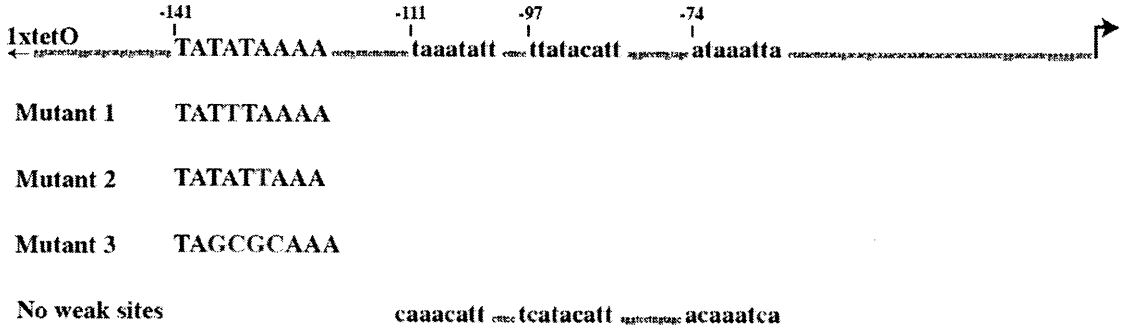


Figure 3.6. Design of the TATA mutants. Shown are the DNA sequence of the *CYC1* core promoter and the mutations (red) used to generate the TATA mutants used in this study.

It is known that a canonical TATA element is related to robust transcription mediated by rapid reinitiation (Yean *et al.*, 1997), which can cause bursting. Indeed, mutating the canonical TATA significantly reduces gene expression noise (Raser and O'Shea, 2004; Blake *et al.*, 2006). The kinetics associated with the non-canonical TATA elements, however, is not known. Moreover, it is unclear how gene-specific activators differentially modulate the utility of the TATA elements. To determine how the distinct TATA elements influence the kinetic roles of activators, we created two types of TATA mutants within the *CYC1* core promoter (Figure 3.6). In Type I mutants, we modified the canonical TATA element with mild (TATA mutant 1), medium (TATA mutant 2) and severe (TATA mutant 3) mutations. In Type II mutants, we eliminated all known non-canonical TATA elements (TATA no weak sites). Then, we titrated tTA activator and measured the resulting mean expression profiles (Figure 3.7) and burst statistics (Figure 3.8) for each of these TATA mutants.

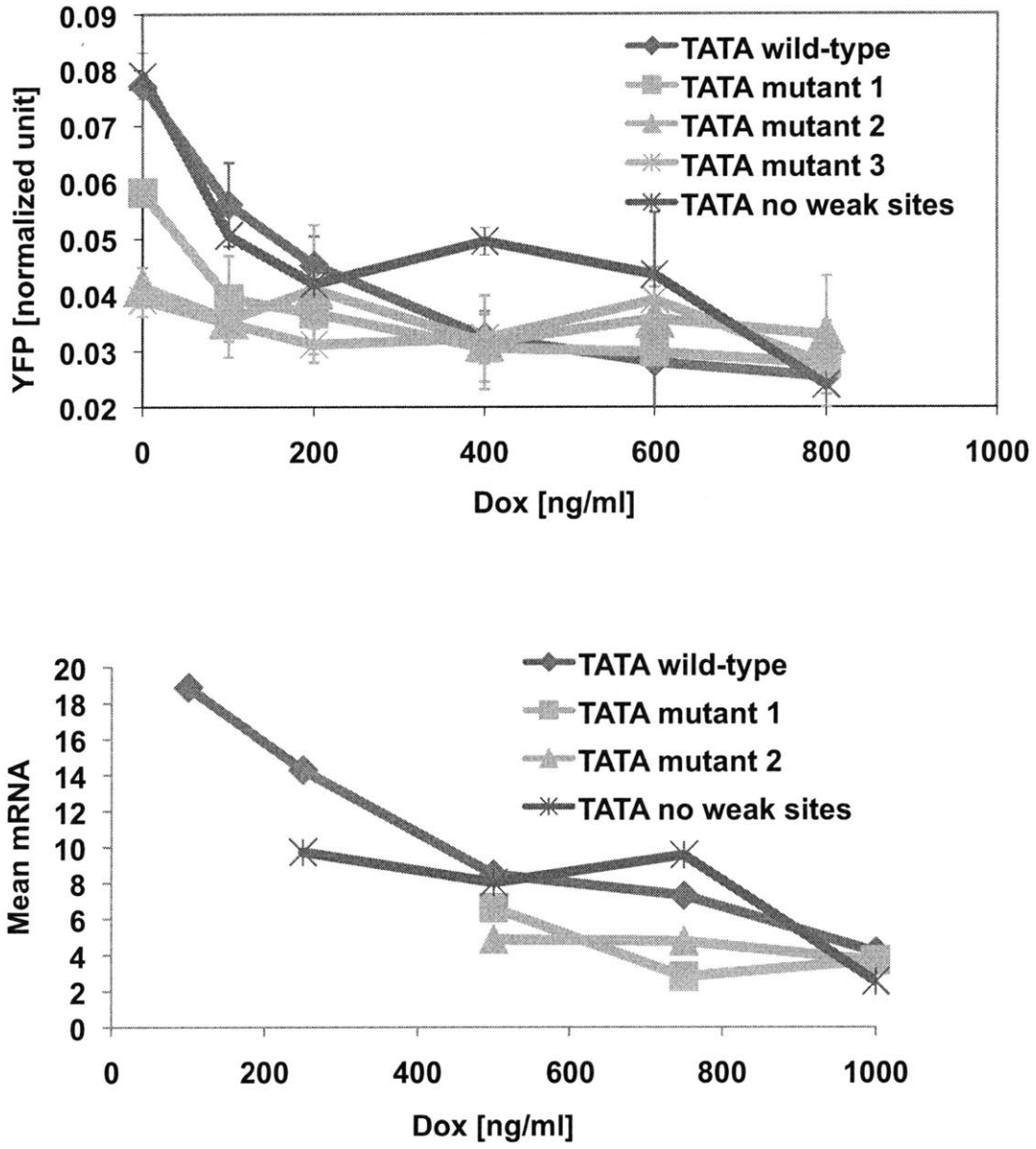


Figure 3.7. The dose response curves of the TATA variants. Shown are the mean YFP expression (top) and mean mRNA expression (bottom) of the TATA variants at different levels of doxycycline. The Type I mutants (TATA Mutants 1 and 2) are colored orange whereas the Type II mutant (TATA no weak sites) is colored purple. All TATA variants are integrated at the *his3* locus.

Both types of TATA mutants are capable of increasing gene expression in response to activators. However, burst statistics, and thereby the gene expression noise, is vastly different between the two types of TATA mutants. Activators predominately regulate burst frequency in the Type I mutants. The burst size largely remains at the basal level in these mutants and only increases slightly at very high activator levels (data not shown). In contrast, activators effectively regulate burst size in the Type II mutant. However, although the Type II mutant exhibits a burst frequency similar to that of the wild-type promoter at very high activator levels (data not shown), it requires a higher activator threshold for full burst frequency modulation. Our results show that both types of TATA elements are capable of supporting initiation independently. The reduced ability of the Type II mutant in regulating initiation at intermediate levels of activators suggests that activators may opt to utilize the non-canonical TATA elements to further increase the rate of initiation. This result implies non-redundant functions for the non-canonical TATA elements. In contrast, only the canonical TATA element can facilitate rapid reinitiation, suggesting that a long TBP residence time at the promoter is crucial to the stability of the reinitiation scaffold.

Interestingly, the Type II mutant has higher burst frequency and mean expression than the wild-type promoter at intermediate activator levels (Figure 3.7). An interpretation of this result is that the non-canonical TATA elements can compete with the canonical TATA element and lead to non-productive initiation. This competition inhibits initiation from both types of TATA elements and thereby reduces the overall burst frequency. Likewise, the Type I mutants allow more efficient initiation at the expense of a lower reinitiation rate, suggesting that the canonical TATA element can

inhibit initiation events from the non-canonical sites and reduce the overall burst frequency. Taken together, our results support the idea that competition exists between the initiation and reinitiation pathways (Huisinga and Pugh, 2007).

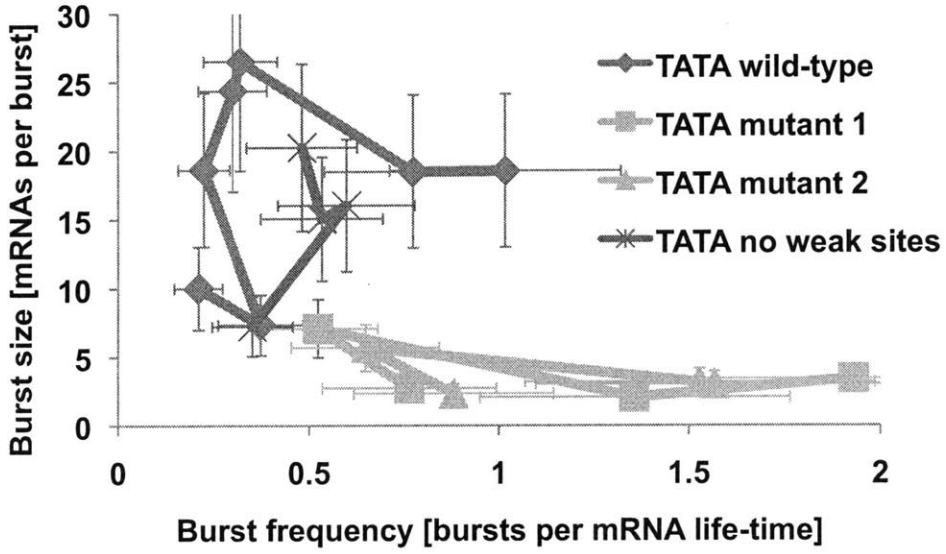


Figure 3.8. Burst frequency and burst size are differentially controlled by activators in TATA mutants. Activators regulate predominately burst frequency in the Type I mutants (orange) and largely burst size in the Type II mutant (purple). The burst frequency and burst size were inferred for each titration point in Figure 3.7 by fitting the steady-state mRNA distribution to the 2-state stochastic model. Error bars represent the estimated technical and biological errors.

The kinetic roles of the distinct TATA elements allow us to develop a molecular model to explain how activators differentially modulate burst frequency and burst size (Figure 3.9). At low activator level, RNA Pol II recruitment could be rate-limiting. Since both TBP and RNA Pol II appear to be always present at the canonical TATA

element of the *CYC1* core promoter (Martens *et al.*, 2001), activators likely regulate initiation through RNA Pol II recruitment to the non-canonical elements.

At moderate activator levels, there is increased occupancy of the TATA elements. Higher levels of overall transcription are associated with increased utilization of both types of TATA elements. At a certain activator level, initiation at the canonical TATA element exceeds a threshold that depends on the inherent stability of the complex assembled. There is a shift to predominant canonical TATA utilization. The reinitiation rate and thereby the burst size increase as the canonical TATA element is preferentially utilized. Here, activators function primarily through the recruitment of Mediator, which increases the stability of reinitiation complex. In this regime, robust reinitiation from the canonical TATA may suppress initiations from the non-canonical sites and reduce the overall burst frequency.

At high activator levels the reinitiation rate, which is limited by the inherent stability of the reinitiation scaffold, attains a maximum. Further addition of activators increases the initiation rate from the canonical TATA element through elevated recruitment of RNA Pol II or downstream factors such as the Mediator and SAGA complexes. Therefore, burst frequency regulation dominates in this regime. The reduced ability of the Type II TATA mutant to support high initiation rates suggests that non-canonical TATA elements are utilized at high activator levels. However, if the model were true, the overall burst size would decrease at high activator levels since initiation events from the non-canonical elements have a low burst size. This seems to be the case for the wild-type 1xtetO promoter at *his3*. To confirm that both types of TATA elements are utilized at high activator levels, we can fit the mRNA distributions to a mixed

negative binomial distribution describing independent bursting from the two different types of TATA elements.

Biochemical assays are required to confirm the detailed molecular mechanism by which a strong acidic activator dictates the pathways to initiation and reinitiation. For instance, to show that activators utilize different TATA elements at different levels, we can perform high resolution ChIP on TBP to estimate the occupancy of different TATA sites. Additional biochemical tests can be performed to quantify the percentage of transcripts initiating at different transcriptional start sites (TSS's), since initiation from different TATA elements appears to result in different TSSs for the *CYC1* core promoter (Li and Sherman, 91). Such experiments will be the subject of future research.

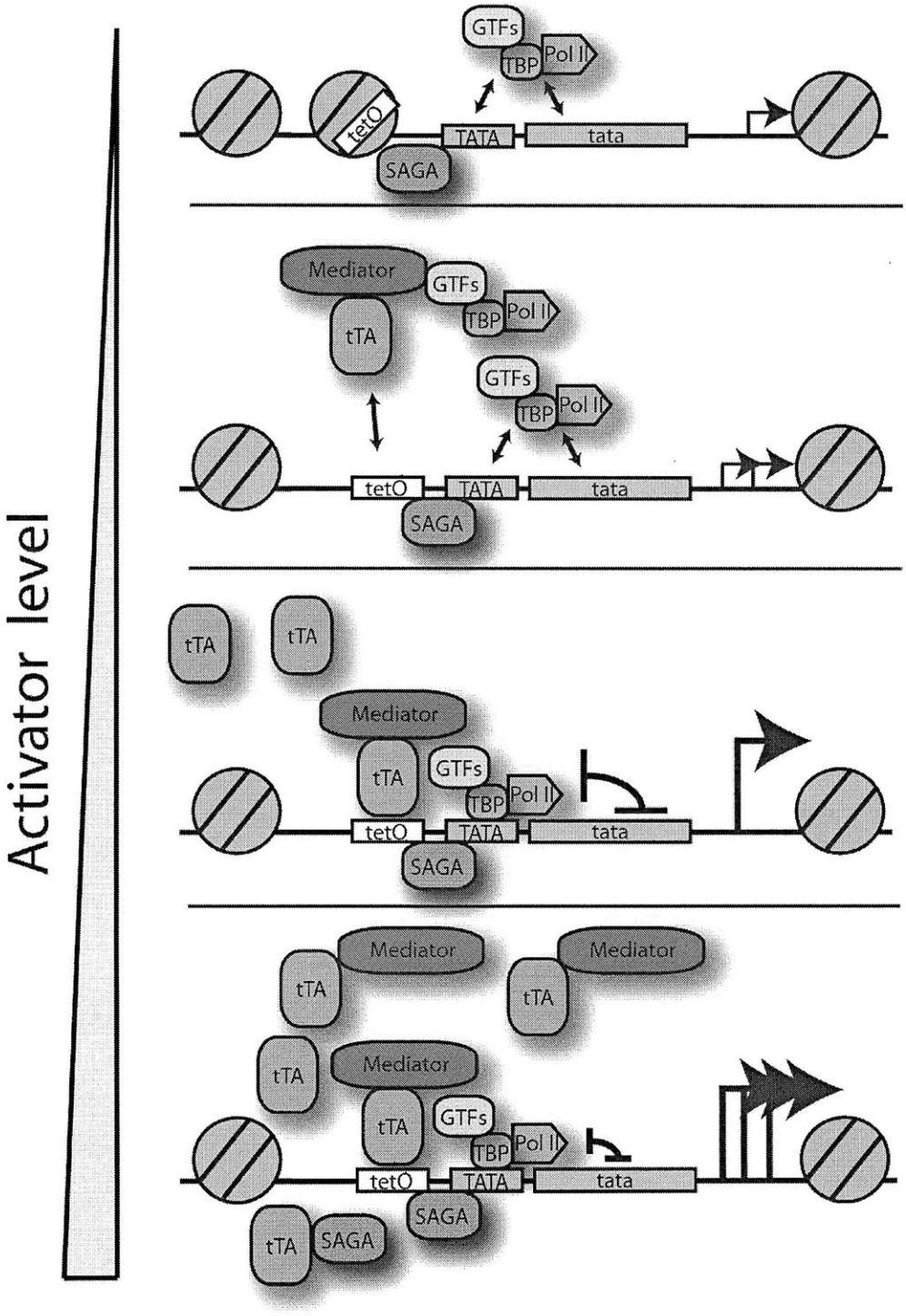


Figure 3.9. A molecular model for the kinetic roles of activators. Utility of distinct TATA elements determines how activators regulate burst size and burst frequency.

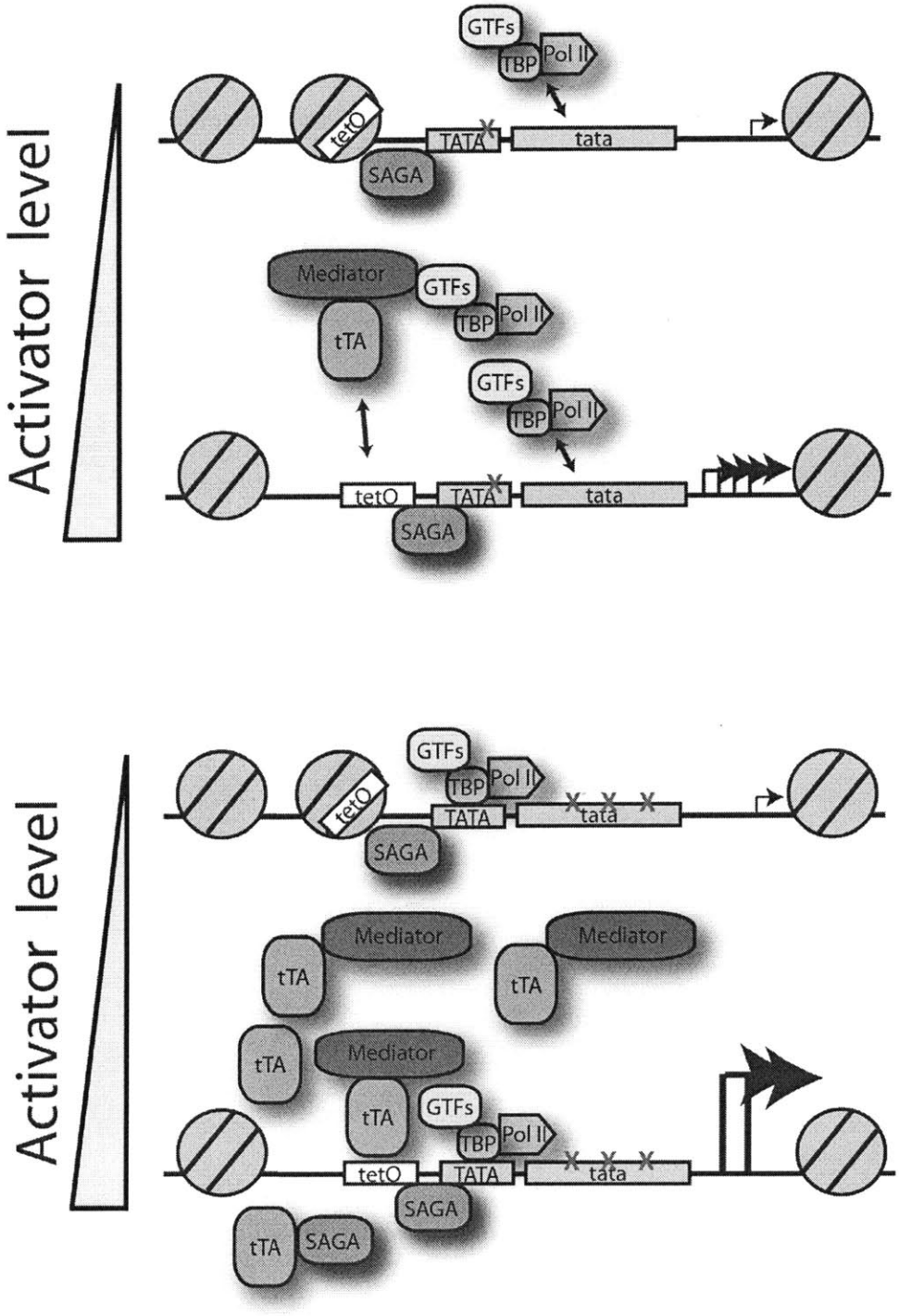


Figure 3.10. A molecular model for the initiation pathways of distinct TATA elements. TATA mutations allow decoupling of initiation pathways and thereby constrain the kinetic role of activators.

3.3.2 Strong activator binding is required for burst size regulation

An important finding in Section 3.3.1 is that a canonical TATA element is required for burst size regulation. This implies that the residence time of TBP at the promoter contributes to the overall stability of the reinitiation complex. This result led us to hypothesize a long residence time at the promoter is also required for the activator to attain a high burst size, since the reinitiation complex also contains the activator according to an *in vitro* biochemical study (Yudkovsky *et al.*, 2000). To test this hypothesis, we decreased the residence time of the activator by creating five tetO mutants based on the PSAM in Section 3.2.5 (Figure 3.11).

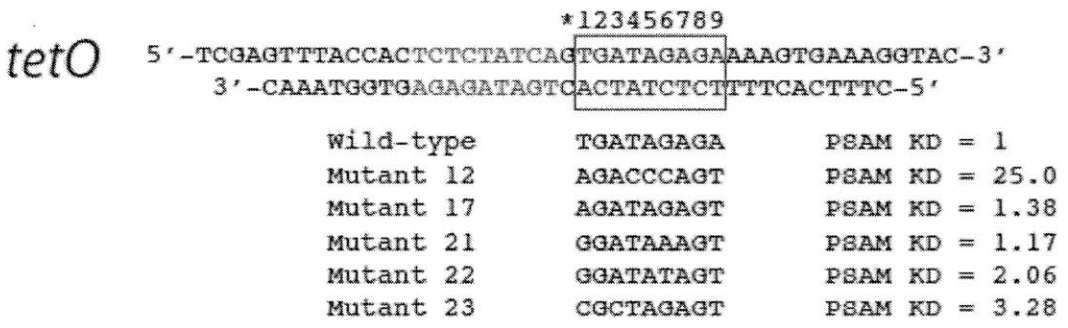
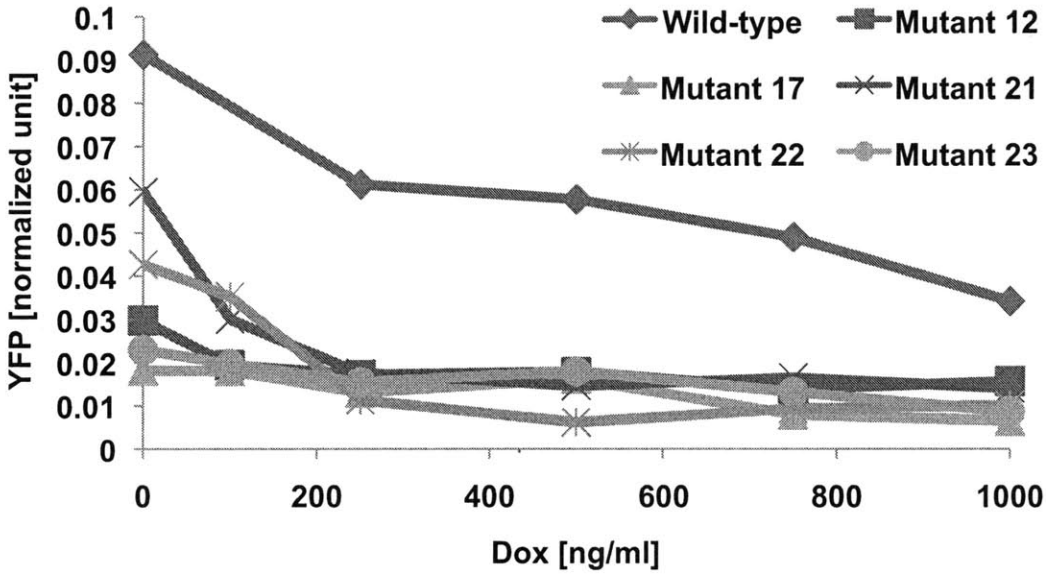


Figure 3.11. Design of the tetO mutants. Shown are the DNA sequence of the wild-type tetO operator and the mutations used to generate the tetO mutants. The dissociation constants (K_D) are predicted using the PSAM and an additive model (Section 3.2.5).

We titrated tTA activators using doxycycline and measured the mean expression profiles (Figure 3.12) of these tetO mutants. Except for Mutant 17, the measured expression profiles generally agree with the K_D values predicted by PSAM. All tetO mutants are capable of increasing gene expression in response to activators. The burst

statistics of the tetO mutants reveal that burst size regulation requires the activator to bind the promoter with a high affinity (Figure 3.13). Although burst frequency is also reduced by a lower binding affinity, the effect is less profound. For example, while burst size regulation is abolished in the weakest mutants (Mutants 12 and 17), activators can still effectively increase the burst frequency, possibly through the utilization of non-canonical TATA mutants. These results further demonstrate that promoter architecture is sufficient to determine the kinetic roles of activators.



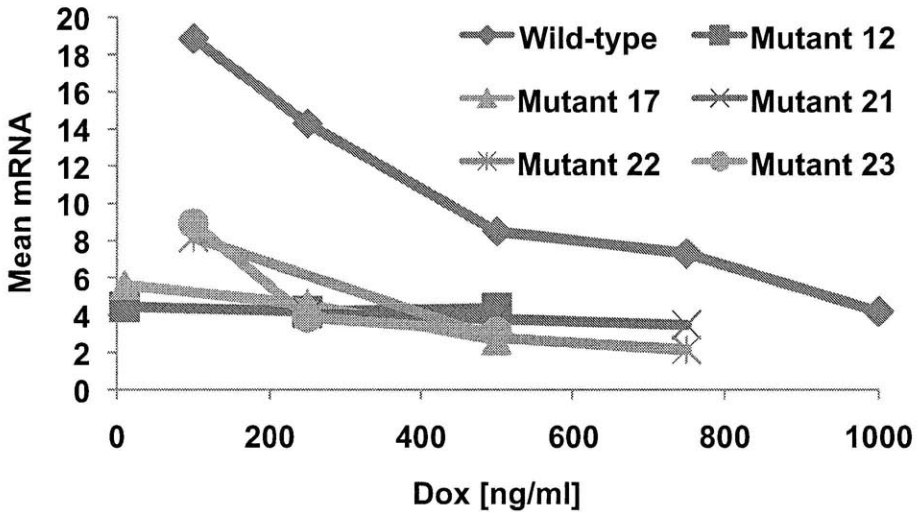


Figure 3.12. The dose response curves of the tetO variants. Shown are the mean YFP expression (top) and mean mRNA expression (bottom) of the tetO variants at different levels of doxycycline. All tetO variants are integrated in the *his3* locus.

While the activator titrations indicate that binding site affinity determines the burst size, the pathway by which strong activator binding supports reinitiation is not obvious. The simplest explanation is that a long activator residence time at the promoter is required for the formation of a stable reinitiation complex, which can then sustain multiple rounds of rapid reinitiation. Multiple mechanisms exist (Figure 3.14). One possible mechanism is that a certain activator residence time is required for covalent modifications of proteins within the reinitiation complex to maintain stability. An alternative mechanism is that longer activator residence times result in a particular form of PIC which will become stable reinitiation complex, as certain forms of PIC contain additional GTFs or coactivators responsible for stabilizing the reinitiation complex. A third possibility is that the initiating activator remains in the reinitiation complex during

the burst. The additional affinity from the binding site increases the stability of the reinitiation complex through cooperative interactions. However, this model is less likely to be correct given the recent finding that gene-specific activators are highly mobile *in vivo* (Karpova *et al.* 2008). To distinguish these possibilities, more detailed mechanistic models are required to make explicit connection between physical events (e.g. activator binding and unbinding) and phenomenological results (e.g. steady-state fluctuations). Previous modeling studies (Blake *et al.*, 2006; Murphy *et al.*, 2007; Sanchez *et al.*, unpublished results) have demonstrated that longer residence times of TBP and activators at the promoter lead to higher gene expression noise. These mechanistic models can be further extended by the incorporation of multiple TATA elements, non-productive pathways, and refractory periods.

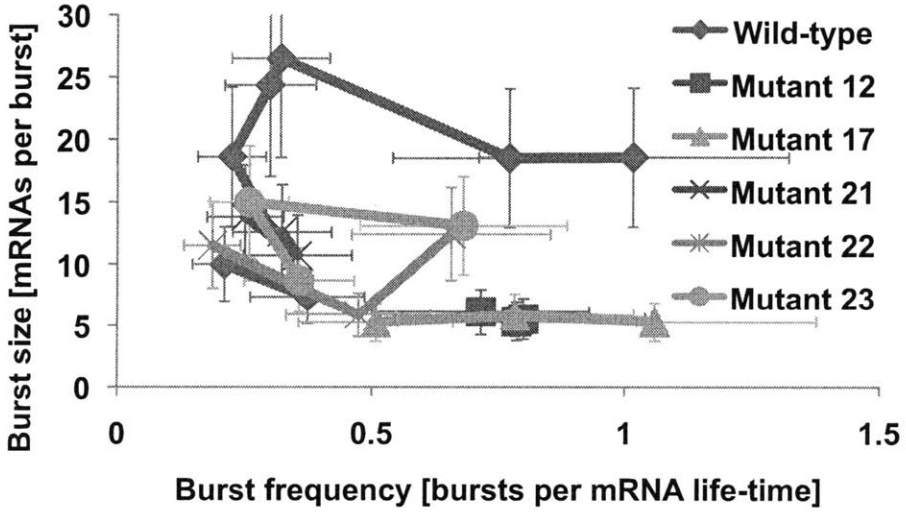


Figure 3.13. Burst size regulation by activators requires a strong activator binding site. The burst frequency and burst size were inferred for each titration point in Figure 3.12 by fitting the steady-state mRNA distribution to the 2-state stochastic model. Error bars represent the estimated technical and biological errors.

3.4 Discussion

While a good deal is known about how activators affect the mean level of gene expression, less is known about how activators affect the kinetic pathways underlying expression heterogeneity. Understanding the kinetic roles of activators allows us to gain insights into how a cell determines the utility of a particular kinetic pathway to fulfill functional needs. This work demonstrates that gene-specific activators can act during both the initiation and reinitiation steps depending on activator levels. Whether activators generally affect both rates of initiation and reinitiation remains an open question, but likely depends on the promoter architecture and the activator type. The ability of an activator to stimulate both slow and fast steps may allow robust and efficient gene activation via kinetic synergism (Herschlag and Johnson, 1993). In addition, such ability allows the same activator to control multiple genes in a kinetically distinct manner.

Although the proposed molecular model (Figures 3.9 and 3.10) is consistent with experimental observations, direct biochemical evidence is lacking. In order to determine the major steps in the assembly process and the binding rates involved, our approach is best supplemented with biochemical assays such as protein-DNA crosslinking technologies with high temporal and spatial resolutions. A kinetic ChIP experiment with high temporal (< 10 minutes) and spatial (< 20 bp) resolutions will allow us to confirm the differential utility of the multiple TATA elements over the course of a burst. A perhaps more promising but less accessible method is to use highly sensitive microscopy or spectroscopy to examine the recruitment of various factors in real-time.

The importance of residence time has broad implications for the stochastic assembly process of transcription machinery. Many components in the transcription machinery, such as the activators and TBP, exhibit highly dynamic interactions with their promoter. Quantitative FRAP experiments have shed light on the dynamics of transcription machinery assembly (Hager *et al.*, 2009). Extensive evidence exists for transient interactions between transcription factors and their chromatin target sites. In particular, different components of the pre-initiation complex appear to arrive at the promoter at different times and only sporadically form a complete functional complex (Sprouse *et al.*, 2008). If this view is indeed true, the residence times of individual components are important since they determine the likelihood of complete assembly. It will be interesting to see how stochastic and reversible protein binding may lead to a fundamental tradeoff between specificity and efficiency of the transcription machinery. Perhaps our finding is related to a recent study (Luijsterburg *et al.*, 2010) where the recognition of DNA lesions by the mammalian nucleotide excision DNA repair (NER) machinery arises from stochastic assembly of multiple NER proteins at the chromatin template. This is an example of kinetic proofreading (Hopfield, 1974), which greatly increases the specificity of the overall assembly process.

Importantly, we show that distinct kinetic pathways can be encoded by a short regulatory region of TATA elements whose mode of regulation is determined by a single type of activator. By configuring the TATA elements, we can specify the utility of these pathways (Figure 3.10). Our findings also illustrate how small changes at TATA-containing promoters may lead to vastly different transcriptional responses and noise levels, highlighting the molecular origin of evolvability of gene expression. This study

provides a plausible mechanism to explain the finding of Landry *et al,* (2007), which demonstrates that sensitivity of gene expression to mutations increases significantly with the presence of a TATA box.

This work shows how various *cis*–factors influence the distribution of regulation between the initiation and reinitiation pathways. Importantly, distributed control of kinetic processes enables the cell to control gene expression heterogeneity independent of expression level. Therefore, our findings have implication for the regulation of various inducible genes that are known to be epigenetically regulated (Octavio *et al.*, 2009). Understanding how activators regulate promoter transitions will allow us to engineer strains with well-defined levels of expression heterogeneity, which can have consequences on phenotypic and population-level fitness. In Chapter 4, we link cell physiology and transcriptional dynamics by showing how a small change in burst statistics can lead to a qualitatively different response in a transcriptional positive feedback loop.

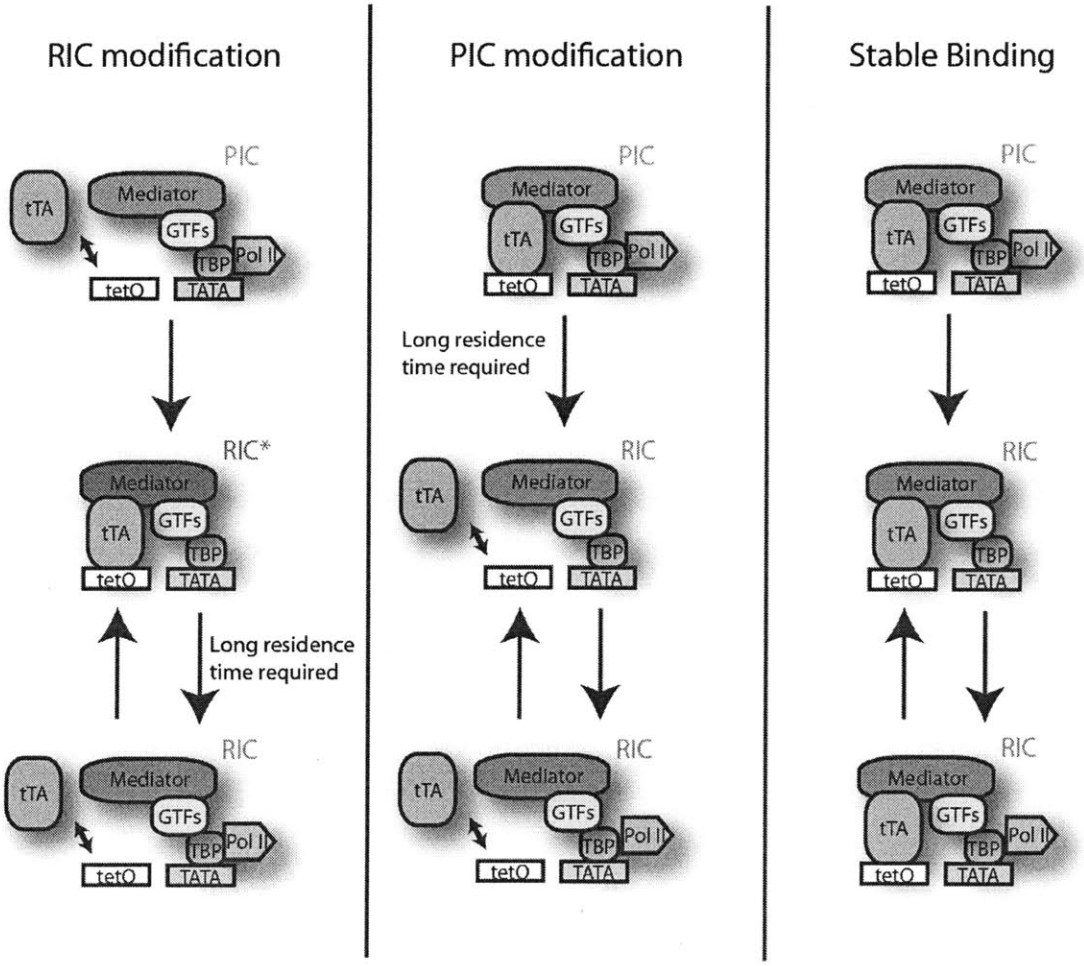


Figure 3.14. Possible mechanisms by which activator residence time influences the stability of the reinitiation complex.

3.5 References

Blake WJ, *et al* (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* 24: 853-865.

Friedman N, Cai L & Xie XS (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* 97: 168302.

Fuda NJ, Ardehali MB & Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461: 186-192.

Gari E, Piedrafita L, Aldea M & Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in saccharomyces cerevisiae. *Yeast* 13: 837-848.

Gossen M, *et al* (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* 268: 1766-1769.

Hager GL, McNally JG & Misteli T (2009) Transcription dynamics. *Mol Cell* 35: 741-753.

Hahn S, Hoar ET & Guarente L (1985) Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 82: 8562-8566.

Herschlag D & Johnson FB (1993) Synergism in transcriptional activation: A kinetic view. *Genes Dev* 7: 173-179.

Hopfield JJ (1974) Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A* 71: 4135-4139.

Huisinga KL & Pugh BF (2007) A TATA binding protein regulatory network that governs transcription complex assembly. *Genome Biol* 8: R46.

Iyer V & Struhl K (1995) Mechanism of differential utilization of the his3 TR and TC TATA elements. *Mol Cell Biol* 15: 7059-7066.

Karmakar R & Bose I (2004) Graded and binary responses in stochastic gene expression. *Phys Biol* 1: 197-204.

Karpova TS, *et al* (2008) Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science* 319: 466-469.

Landry CR, Lemos B, Rifkin SA, Dickinson WJ & Hartl DL (2007) Genetic properties influencing the evolvability of gene expression. *Science* 317: 118-121.

Li WZ & Sherman F (1991) Two types of TATA elements for the CYC1 gene of the yeast saccharomyces cerevisiae. *Mol Cell Biol* 11: 666-676.

Luijsterburg MS, *et al* (2010) Stochastic and reversible assembly of a multiprotein DNA repair complex ensures accurate target site recognition and efficient repair. *J Cell Biol* 189: 445-463.

Maamar H, Raj A & Dubnau D (2007) Noise in gene expression determines cell fate in bacillus subtilis. *Science* 317: 526-529.

Martens C, Krett B & Laybourn PJ (2001) RNA polymerase II and TBP occupy the repressed CYC1 promoter. *Mol Microbiol* 40: 1009-1019.

Munsky B & Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* 124: 044104.

Murphy KF, Balazsi G & Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A* 104: 12726-12731.

Octavio LM, Gedeon K & Maheshri N (2009) Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. *PLoS Genet* 5: e1000673.

Ptashne M & Gann A (1997) Transcriptional activation by recruitment. *Nature* 386: 569-577.

Pugh BF (2000) Control of gene expression through regulation of the TATA-binding protein. *Gene* 255: 1-14.

Raj A, Peskin CS, Tranchina D, Vargas DY & Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4:

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A & Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 5: 877-879.

Raj A & van Oudenaarden A (2009) Single-molecule approaches to stochastic gene expression. *Annu Rev Biophys* 38: 255-270.

Raser JM & O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304: 1811-1814.

Roeder RG (2005) Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* 579: 909-915.

Shahrezaei V & Swain PS (2008) Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci U S A* 105: 17256-17261.

Sizemore C, Wissmann A, Gulland U & Hillen W (1990) Quantitative analysis of Tn10 tet repressor binding to a complete set of tet operator mutants. *Nucleic Acids Res* 18: 2875-2880.

Sprouse RO, *et al* (2008) Regulation of TATA-binding protein dynamics in living yeast cells. *Proc Natl Acad Sci U S A* 105: 13304-13308.

Struhl K (1986) Constitutive and inducible saccharomyces cerevisiae promoters: Evidence for two distinct molecular mechanisms. *Mol Cell Biol* 6: 3847-3853.

To TL & Maheshri N (2010) Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* 327: 1142-1145.

Weake VM & Workman JL (2010) Inducible gene expression: Diverse regulatory mechanisms. *Nat Rev Genet* 11: 426-437.

Yean D & Gralla J (1997) Transcription reinitiation rate: A special role for the TATA box. *Mol Cell Biol* 17: 3809-3816.

Yudkovsky N, Ranish JA & Hahn S (2000) A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408: 225-229.

4 Transcriptional bursting can lead to bimodal gene expression in positive transcriptional feedback loops without bistability

4.1 Introduction

When a cell must unambiguously commit to a particular gene expression program, often a digital change occurs in a key regulator's expression (Alon, 2006). Decision-making circuitry within metabolic (Acar *et al.*, 2005), developmental (Xiong and Ferrell, 2003), and synthetic gene regulatory networks (Ingolia and Murray, 2007; Issacs *et al.*, 2003; Ajo-Franklin *et al.*, 2007) uses positive transcriptional feedback loops to provide bimodal, all-or-none expression of a regulator.

In all these cases, the promoter response in the absence of feedback was either measured or assumed to be sigmoidal. For example, in synthetic regulatory constructs where a TF regulates its own expression, a Hill-type equation is often used to model the open-loop response. Open-loop responses with Hill coefficients greater than 1 are often attributed to direct cooperative binding of TFs to promoters, or indirect cooperativity via nucleosome displacement (Miller and Widom, 2003). The resulting sigmoidal response is a necessary system property for the existence of bistability in deterministic theoretical models (see Section 4.2.2 for details).

Meanwhile, it has become increasingly appreciated that stochastic noise in gene expression can alter the qualitative behavior of transcriptional regulatory networks. Recent single molecule approaches have revealed that gene expression often occurs in

bursts of transcription and/or translation (Maheshri and O'Shea, 2007; Raj and van Oudenaarden, 2008). The burst statistics can be characterized by two parameters: the burst size (number of mRNA or proteins produced per transcriptional activation event) and the burst frequency (number of transcriptional activation events per mRNA or protein lifetime). The mean level of gene expression is set by the product of burst size and burst frequency, but the noise in gene expression is proportional to the burst size (see Chapter 3 for details).

The effect of noise in gene expression on the properties of the positive feedback motif has been studied both theoretically (Kepler and Elston, 2001; Friedman *et al.*, 2001; Karmakar and Bose, 2007). Interestingly, both Friedman *et al.* (2006) and Karmakar and Bose (2007) predict that a bimodal expression distribution can occur with positive feedback even when the open-loop response is linear and graded with a Hill coefficient <= 1 (see Section 4.2.3 for details). This is similar to another theoretical study demonstrating that noise in enzymatic futile cycles can induce bimodal activity when a deterministic description only admits one stationary solution (Samoilov *et al.*, 2005).

Here, we demonstrate experimentally that a transcription factor (TF)-promoter pair with a non-cooperative, graded open-loop response can exhibit an "all-or-none" steady-state bimodal response when reconfigured in a positive feedback motif. As predicted from the theory, a robust bimodal response occurs only when the timescale of transcriptional bursts is less than the lifetime of the TF; in other words, the promoter possesses a low maximum burst frequency but a corresponding burst size large enough to trigger expression. Further stabilization of the TF eliminates the response. However, while a simple stochastic model qualitatively describes these results, it fails to capture the

observed stability of the two states. This prompted a further investigation that revealed the importance of a third process - nuclear transport of the TF. The simple stochastic model compared well to experimental results when the TF was tagged with a strong nuclear localization signal (NLS). Interestingly, a weaker NLS leads to an enhanced bimodal response.

4.2 Theory

4.2.1 A deterministic model of gene regulation

To describe the steady-state, mean level of expression of protein p for open-loop constructs, we employed a Hill-like model to describe the gene regulatory function (Bintu *et al.*, 2005):

$$p = k_{\min} + (k_{\max} - k_{\min}) \frac{(T/K)^n}{(T/K)^n + 1} \quad [4.1]$$

Here k_{min} and k_{max} are production rates of protein expression already normalized by the protein degradation rate, and hence have units of protein concentration or number. In this simple thermodynamic model, the affinity of TF to its binding site and the transcription factor concentration dictate the probability of particular promoter states (bound or unbound TF). The transcription rate from the promoter, and hence the protein expression, is proportional to these states. For example, if the TF is an activator, expression is typically modeled as proportional to the bound TF promoter state.

For our particular system (Tet-OFF, see Section 3.2.1 for details), the dissociation constant (K) between tTA (T) and the promoter can be modulated by addition of doxycycline. The explicit dependence of K on doxycycline levels [dox] is:

[4.2] $$K = K_o(1 + 2[dox]/K_S + ([dox]/K_S)^2)$$

where K_0 is the dissociation constant in the absence of doxycycline and K_S is the dissociation constant between tTA and doxycycline. The quadratic [dox] dependence of K arises from the fact that there are two binding sites for doxycycline on the tTA dimer and that only unbound tTA was active, as even single doxycycline-bound tTA has its DNA binding affinity reduced by 10^3 fold (Henssler *et al.*, 2005). A similar approach has been taken by Murphy *et al.* (2007).

4.2.2 A deterministic model of positive feedback and bistability

The above model can be easily reformulated to describe the steady-state response of a positive transcriptional feedback loop. A definitive feature of the positive feedback loop (Figure 4.1-A) is bistability in its gene expression. This feature can be explained by modeling its dynamics with minimal model that couples transcription and translation.

[4.3] $$\frac{dT}{d\tau} = T_{\min} + (T_{\max} - T_{\min})\frac{T^n}{K^n + T^n} - T$$

T denotes the concentration of transcription factor which activates its own expression. There are two terms for production – basal production (T_{min}) and activator mediated production (the second term of R.H.S.). The degradation (the last term of R.H.S.) is

assumed to be first-order. Although the above equation can be solved to obtain steady-state fixed-points, it is best to be analyzed graphically. The intersection points of production and degradation curves represent the fixed points. The key parameter is n, the Hill coefficient, which signifies the sharpness of response towards activator concentration. For $n = 1$, there can only be one fixed point for any $T_{min} > 0$ (Figure 4.1-B). This fixed point is stable. Shifting the production curve (by changing K, for example) can result in a different fixed point, but cannot switch the stability of the fixed point since there is no bifurcation possible.

Qualitatively different behavior can be observed for $n = 2$. Physically, the sharpness of response increases as n increases. In order words, $n > 1$ signifies cooperativity in production. The outcome for $n > 1$ is that the production curve becomes sigmoidal and ultrasensitive. For $n = 2$, the production curve can intersect the degradation curve three times (Figure 4.1-C) to give 3 fixed points. However, only the leftmost and rightmost fixed points are stable. Notice that as K gets larger, the production curve shifts in a way such that it only intersects with the degradation curve once. When the production curve becomes tangential to the degradation curve, the unstable and one of the stable fixed points collide and annihilate, resulting in a saddle node bifurcation. To predict the behavior of the system for different values of the parameters K and T_{min}, we construct the analytical bifurcation diagram to illustrate where bifurcation occurs for $n = 2$ (Figure 4.1-D). Bistability can be attained under a fairly wide range of parameter values. In summary, cooperativity ($n > 1$) is necessary (though not sufficient) for bistability, as it is necessary for the production curve to intersect with the degradation curve at 3 points.

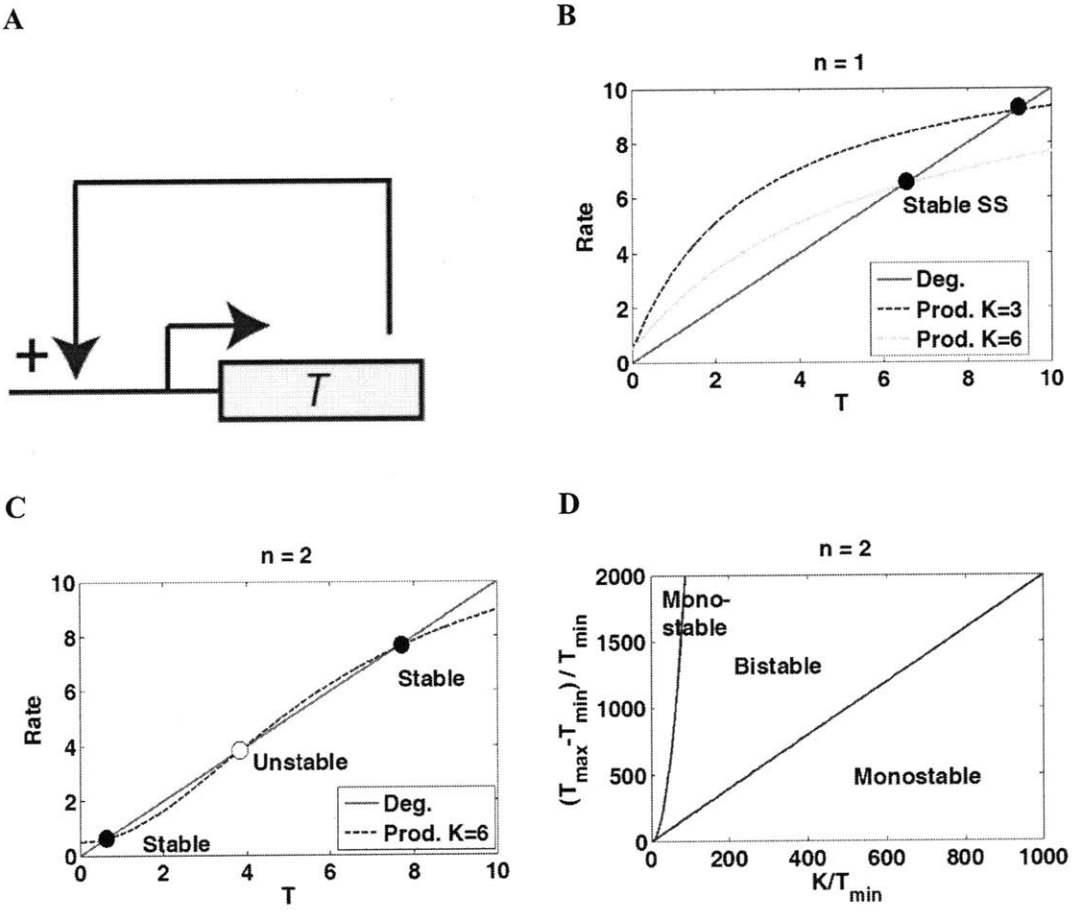


Figure 4.1. Positive feedback and the emergence of bistability. (A) An autocatalytic gene in which the transcription factor (T) activates its own expression. (B) In the absence of cooperativity ($n = 1$) the system can only admit 1 stable steady state. Other parameters are $T_{min} = 0.5$ and $T_{max} = 12$. (C) Bistability is possible with cooperativity ($n = 2$). Other parameters are $T_{min} = 0.5$ and $T_{max} = 12$. (D) Bifurcation diagram illustrates that bistability is achievable under a wide range of parameters.

4.2.3 Friedman model for positive feedback

When burst frequency is regulated by TF, (Friedman *et al.*, 2006) showed that the steady-state protein distribution in the presence of transcriptional position feedback loop is:

$$P(x) \propto (x/K)^{\tilde{\lambda}_0 C} \exp(-\frac{\tilde{\lambda}_0 (x/K)}{R})(1+K/x)^{\tilde{\lambda}_0} \quad [4.4]$$

where we have reparameterized their results with $C = 1 - 1/\tilde{\lambda}_o + \varepsilon$ and $R = \tilde{\lambda}_o \mu/\gamma K$.

The qualitative effect of the burst size and burst frequency on the occurrence of bimodality can be conveniently understood by considering equation [4.4] and its corresponding phase plot (Figure 4.2). C (y-axis of the phase plot) is determined by the maximum burst frequency. R (x-axis of the phase plot) is a measure of the maximum expression level relative to *K*. When expression levels are larger than *K*, the linear promoter response begins to saturate. Frequent bursts will always lead to some non-zero level of activation with basal expression. If $C \geq 1$ (region I), a unimodal, non-zero expression level is always observed. The mean expression level can be increased independently of C by changing R, either by increasing the burst size, or by decreasing *K*, thereby increasing the actual (not maximum) burst frequency. For $C < 1$ the activator degradation rate is larger than the maximum promoter activation rate and bimodal expression is possible. At low levels of R, expression is peaked at zero (region II), but the distribution gets a longer tail as R increases. Finally, upon entering region III, the distribution becomes bimodal. Although formally the model continues to predict a bimodal distribution for larger values of R, the actual percentage of OFF cells drops to

vanishingly small values. Region IIIb demarcates a region where the percentage of OFF cells is < 5% of the total population, to provide a sense of the bimodal region.

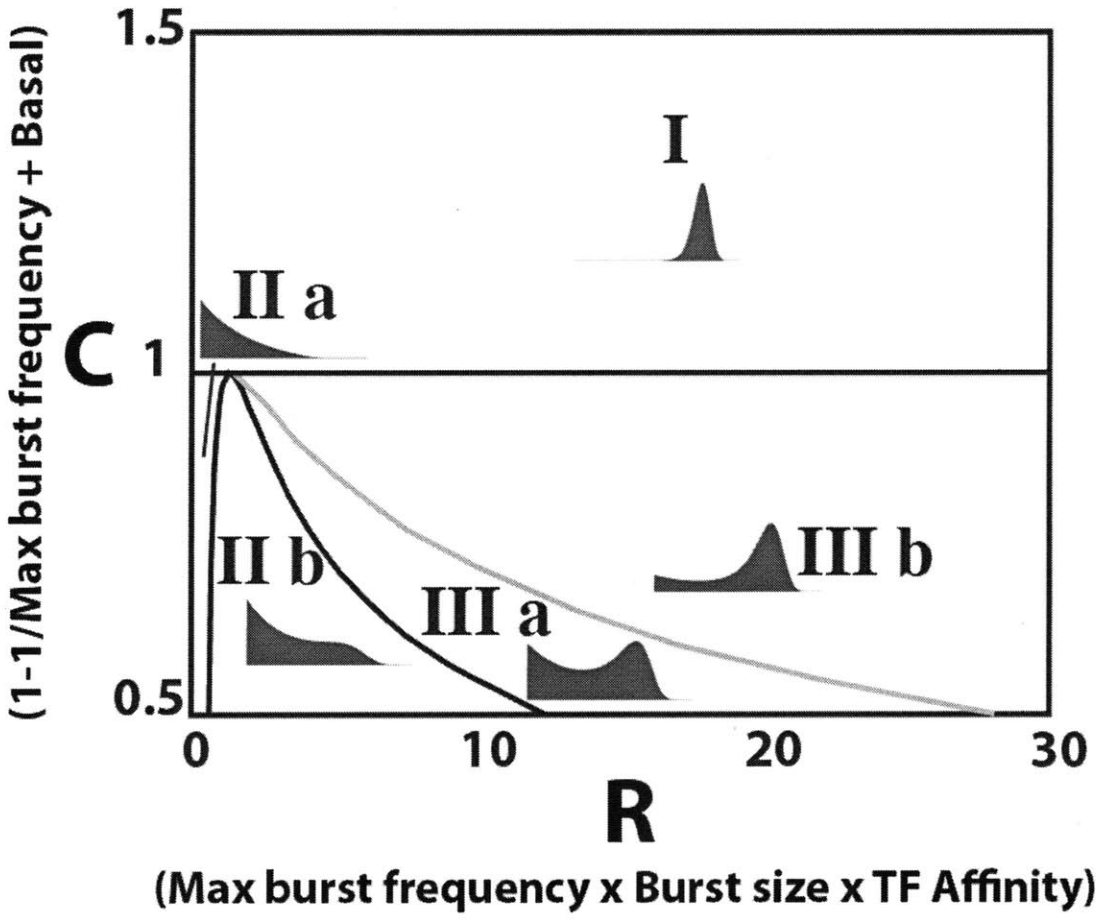


Figure 4.2. A phase diagram describing the qualitative behavior of the closed-loop response. Using the Friedman model (Section 4.2.3), the closed-loop response is predicted by parameters characterizing the open-loop promoter response. Bimodal expression occurs when the maximum burst frequency is low and the burst size is large enough (compared to the TF dissociation constant) that the promoter can turn on in positive feedback (Region IIIa).

Figure 4.3 describes the essential physics of how noise and positive feedback could combine to convert a linear open-loop response into an "all-or-none" bimodal closed-loop response. Consider an autoregulatory gene at low levels of expression. Transcriptional bursts of mRNA lead to bursts of activator protein. If the activator is very long-lived, it will feedback on its promoter and the system will proceed to the deterministic steady-state. This corresponds to the case of high burst frequency. If the activator is shorter-lived though, it may degrade before activator levels are high enough to sustain high expression. Rare, infrequent events, made possible by large burst sizes and small burst frequencies, lead to high activator levels and switch the system to high expression levels. Because the activator is still short-lived, infrequently all activators will degrade before the next burst occurs. This switches the system back to low expression levels. Therefore the burst frequency, determined by both promoter firing and activator stability, plays a crucial role in determining whether noise-induced bimodality will occur in positive feedback.

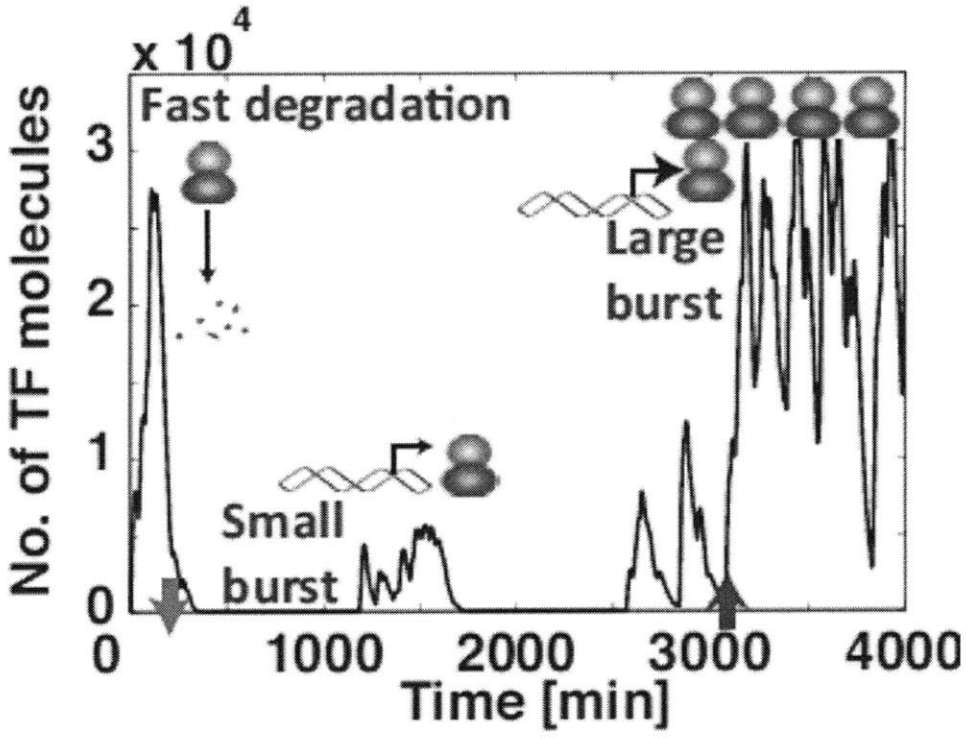


Figure 4.3. A simulated time series describing expression in a bursty positive feedback. Infrequent promoter firing and an unstable activator lead to a bimodal closed-loop response.

4.3 Materials and Methods

4.3.1 Estimation of kinetic parameters for the deterministic model

Using the results of the open-loop experiments, we simultaneously fit 7xtetO and 1xtetO titration data to Equations [4.1] and [4.2], demanding that K_S was the same in both cases. We estimated the ratio T/K_o as we could not estimate T directly. However, because in the most of the open-loop experiments tTA was constitutively expressed from the *ADH1* promoter, we defined 1 effective tTA unit as the tTA level expressed from the *ADH1* promoter in the absence of doxycycline. A $K_S \sim 500$ (ng/mL doxycycline) worked

consistently for all our datasets. We always obtained Hill coefficients very close to 1, which was not surprising as the measured open-loop responses appear non-cooperative (see Section 4.4.1). However, because the response never saturates, it was impossible to accurately estimate k_{max}. Therefore, to confirm the accuracy of equation [4.1] over the full promoter response and to accurately estimate k_{max}, we increased tTA level using the positive feedback loop. We assessed the tTA levels relative to *ADH1* promoter-driven tTA expression (1 effect tTA unit) by employing a tTA-CFP fusion protein and comparing the relative fluorescence levels. The transcriptional activity of tTA was not affected by the CFP tag. However, we note that in the presence of doxycycline, the observed CFP signal of the tTA-CFP fusion does not represent the active tTA level as some fraction is bound by doxycycline. We corrected for this by using equation [4.2] to estimate the active tTA fraction. We then converted the active tTA level in the feedback context into effective tTA units using the CFP signal from tTA-CFP driven by *ADH1* promoter in the absence of doxycycline.

Using the full promoter response obtained in the feedback context, we determined the three macroscopic kinetic parameters for the gene regulatory function – namely k_{min}, k_{max}, and K. The k_{min} was directly obtained from the minimum reporter response for a given promoter in the absence of tTA. To obtain k_{max} and K of the gene regulatory function, we determined the promoter response curves using both protein and mRNA measurements. We fit these curves to Equation [4.1] using MATLAB's nlinfit program (MathWorks) to obtain k_{max} and K (Table 4.1).

4.3.2 Estimation of kinetic parameters for the stochastic model

We fit mRNA FISH data to numerical simulations of the reaction scheme [3.1] in Section 3.2.3, using maximum likelihood estimation (MLE). As in the case described in Section 3.2.4, close examination of the fits showed a ridge of constant likelihood for constant $\tilde{\mu}/\tilde{\gamma}$ (Figure 4.4), implying that a two parameter burst size / burst frequency model was appropriate and activation was rare compared to inactivation. There is uncertainty with respect to how the values of $\tilde{\mu}$ and $\tilde{\gamma}$ change between the 1xtetO and 7xtetO cases because only the burst size can be accurately estimated. We favor a model where it is $\tilde{\gamma}$ that changes when additional tetO binding sites are introduced because a increased number of TF's keeps the promoter in a "active" conformation longer, decreasing the promoter inactivation rate. As such, we held $\tilde{\mu}$ fixed, and used MLE to determine $\tilde{\lambda}$ and $\tilde{\gamma}$ from the experimental data. In doing so, we assume the transcriptional initiation rate in the active state (which $\tilde{\mu}$ represents) is the same for 1xtetO and 7xtetO. We chose a fixed value of 500 for $\tilde{\mu}$ for both 1xtetO and 7xtetO, and obtained values for $\tilde{\lambda}$ and $\tilde{\mu}/\tilde{\gamma}$ at different tTA levels. The value of $\tilde{\mu}$ was estimated to be slightly larger than the maximum number of mRNAs observed in a cell (467, from 7xtetO grown with no doxycycline).

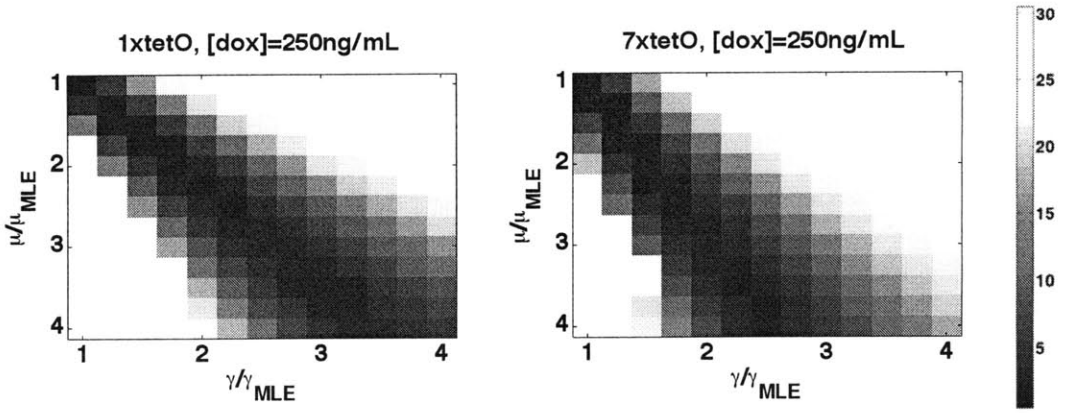


Figure 4.4. The parameters γ and μ cannot be estimated independently. The mRNA distributions obtained by FISH experiments were fit to numerical simulations of the reaction scheme [3.1]). The likelihood values are plotted as a function of $\tilde{\gamma}$ and $\tilde{\mu}$ for two samples in Figure 4.9 in Section 4.4.2. The color bar represents the negative log-likelihood score. The fits show a ridge of constant likelihood for constant $\tilde{\mu}/\tilde{\gamma}$. Similar results were obtained for other samples in Figure 4.9.

4.3.3 Relating the microscopic kinetic parameters to the stochastic model

When we plotted the logarithm of the mRNA noise versus the logarithm of mRNA abundance (Figure 4.12), we observed a slope ~ -0.5, the expected scaling law for the open-loop response in both the presence and absence of feedback. This implies the burst frequency is being regulated by the transcriptional activator, and the intercepts give the mean burst size (Table 4.1). The scaling law has been observed generally in global studies of noise at yeast promoters (Bar-Even *et al.*, 2006; Newman *et al.*, 2006). Therefore, we model activation as:

[4.5]
$$\tilde{\lambda} = \tilde{\lambda}_o(\varepsilon + \frac{\hat{x}}{\hat{x} + \hat{K}})$$

where K denotes the dissociation constant between TF and the promoter, $\tilde{\lambda}_0$ denotes the maximum burst frequency and ε denotes basal bursts in the absence of the TF. The $k_{\max}$ and $k_{\min}$ parameters of the deterministic model can be explicitly related to the parameters of the stochastic model:

$$k_{\min} = \frac{\varepsilon\tilde{\lambda}_o}{\varepsilon\tilde{\lambda}_o + \tilde{\gamma}}\tilde{\mu}$$
$$k_{\max} = \frac{\tilde{\lambda}_o(1+\varepsilon)}{\tilde{\lambda}_o(1+\varepsilon) + \tilde{\gamma}}\tilde{\mu} \quad [4.6]$$

If inactivation rates are *always* higher than activation rates (even with saturating amounts of transcription factor) and $\varepsilon << 1$, then equation [4.7] reduces to

$$k_{\min} = \frac{\varepsilon\tilde{\lambda}_o}{\tilde{\gamma}}\tilde{\mu} = \varepsilon\tilde{\lambda}_o\frac{\tilde{\mu}}{\tilde{\gamma}}$$
$$k_{\max} = \tilde{\lambda}_o\frac{\tilde{\mu}}{\tilde{\gamma}} \quad [4.7]$$

where k_{max} (of mRNA production) is simply the product of the maximum burst frequency times the burst size. If burst size is constant across activator levels for a given promoter (which seems to be the case for tetO, Figure 4.11), $\tilde{\lambda}_0$ and ε can be determined once k_{max} and k_{min} are known and all parameters required for the positive feedback model are defined.

Additional materials and methods can be found in the Appendices.

4.4 Results

4.4.1 A bimodal, "all-or-none" response in a transcriptional positive feedback loop without evidence of bistability

To understand how promoter structure relates to the Hill coefficient, we employed the widely used tet-OFF system, adapted for budding yeast (Gari *et al.*, 1997). The tet-Transcriptional Activator (tTA) binds to a tet operator (tetO) sequence in the absence of doxycycline. We constructed yeast strains without (open-loop) and with feedback (closed-loop) using previously designed promoters with 1 (1xtetO) and 7 (7xtetO) binding sites. With 1xtetO in positive feedback, the reporter exhibits a graded steady-state response to changes in feedback strength, while 7xtetO exhibits a bimodal response (Figure 4.5). One explanation is that 7xtetO has a sigmoidal open-loop response due to cooperative binding of tTA to multiple binding sites, resulting in bistability (Ajo-Franklin *et al.*, 2007; Becskei *et al.* 2001). Yet, if one accounts for the binding of doxycycline to the tTA dimer (Section 4.2.1), both 1xtetO and 7xtetO exhibit a graded open-loop response, with a Hill coefficient of ~1 (Figure 4.6). To eliminate the possibility of altered doxycycline-binding, we directly titrated tTA levels using a galactose-inducible promoter, and confirmed the non-cooperative response (Figure 4.7). In addition, the open-loop response of both promoters in a strain that contains a closed-loop promoter is also non-cooperative (Figure 4.8).

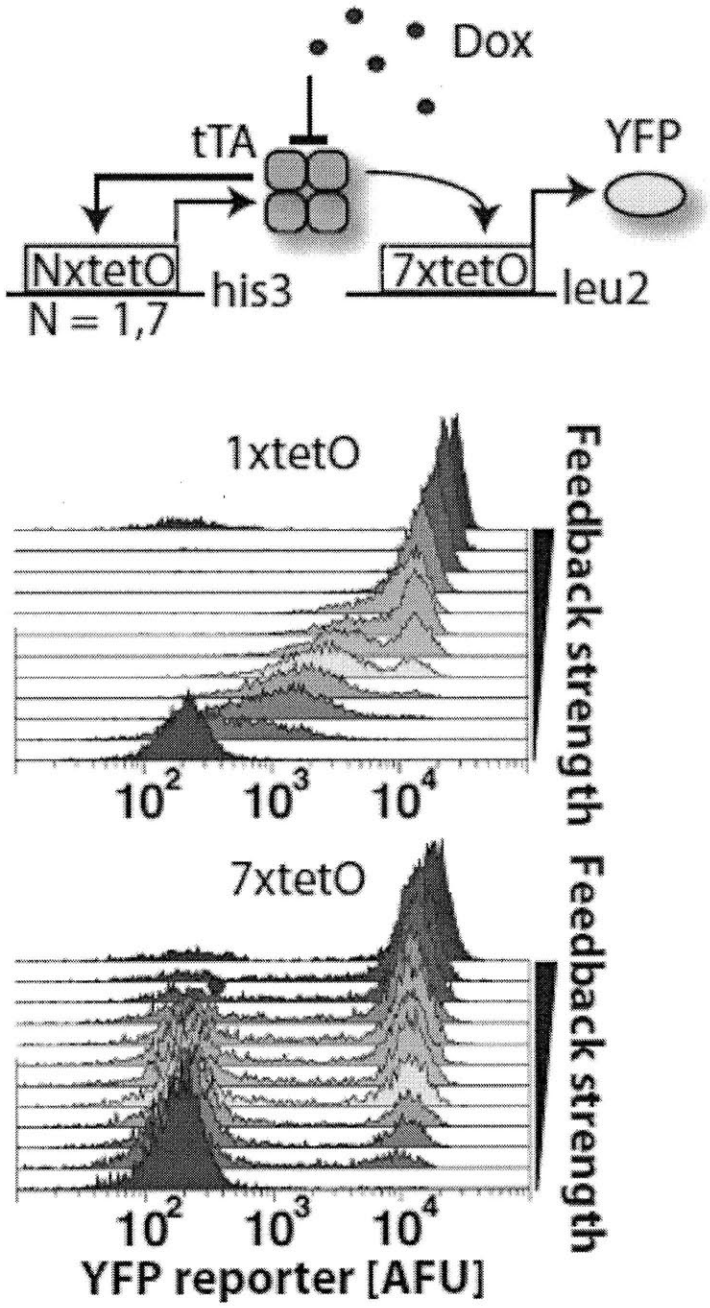


Figure 4.5. Bimodal gene expression in positive feedback with multiple binding sites. Yeast strains engineered with 1xtetO and 7xtetO in a closed loop configuration were grown to steady-state. Feedback strength was modulated by varying doxycycline from 0 to 2000 ng/mL. Resulting steady-state expression distributions were obtained from flow cytometry.

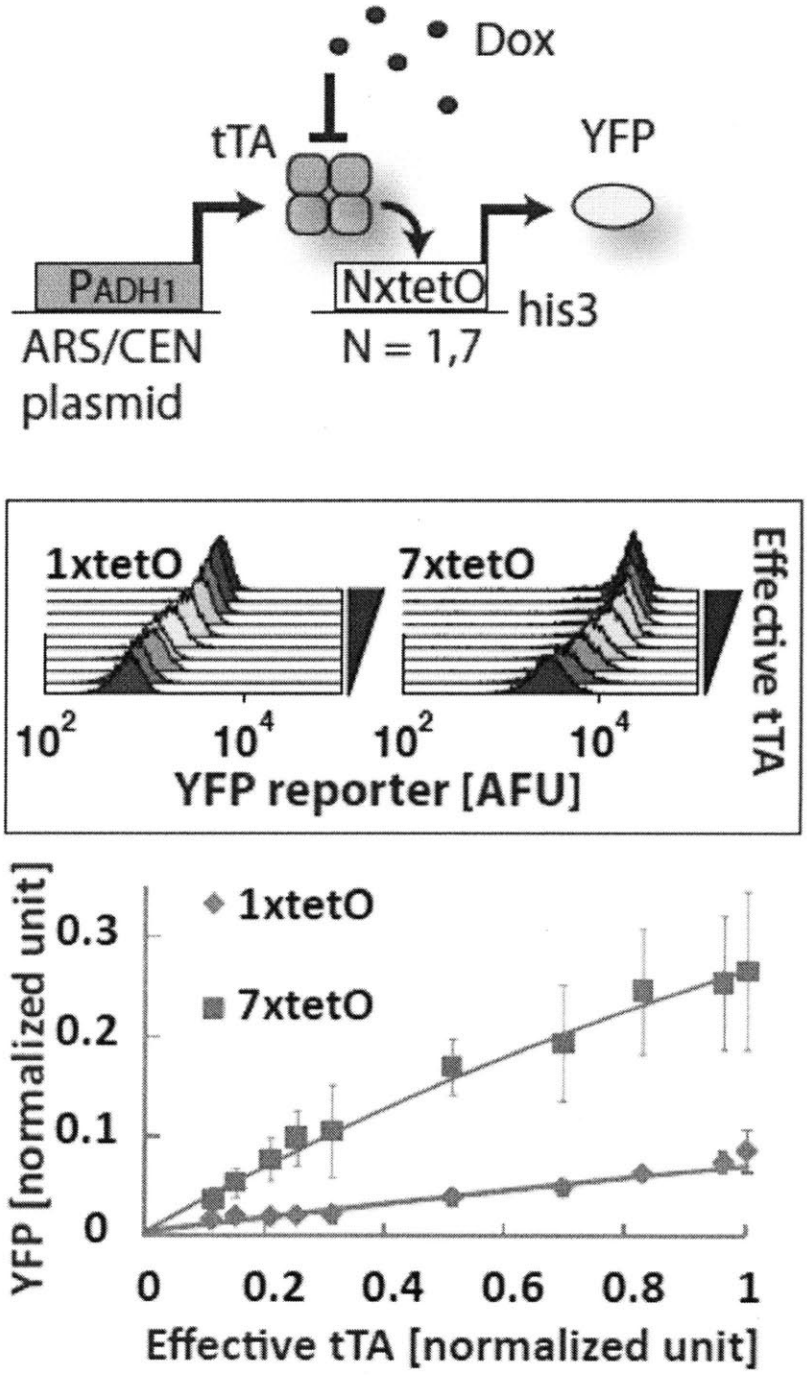


Figure 4.6. The open-loop response of the 1xtetO and 7xtetO promoters. The dose response curves determined by titration with 0 – 2000 ng/mL doxycycline are non-cooperative with a Hill coefficient of ~1. Flow cytometric analysis confirms the response is graded at the single cell level (Inset). The effective tTA level was calculated from

doxycycline concentration using equation [4.2]. Error bars indicate the s.d. of triplicate samples.

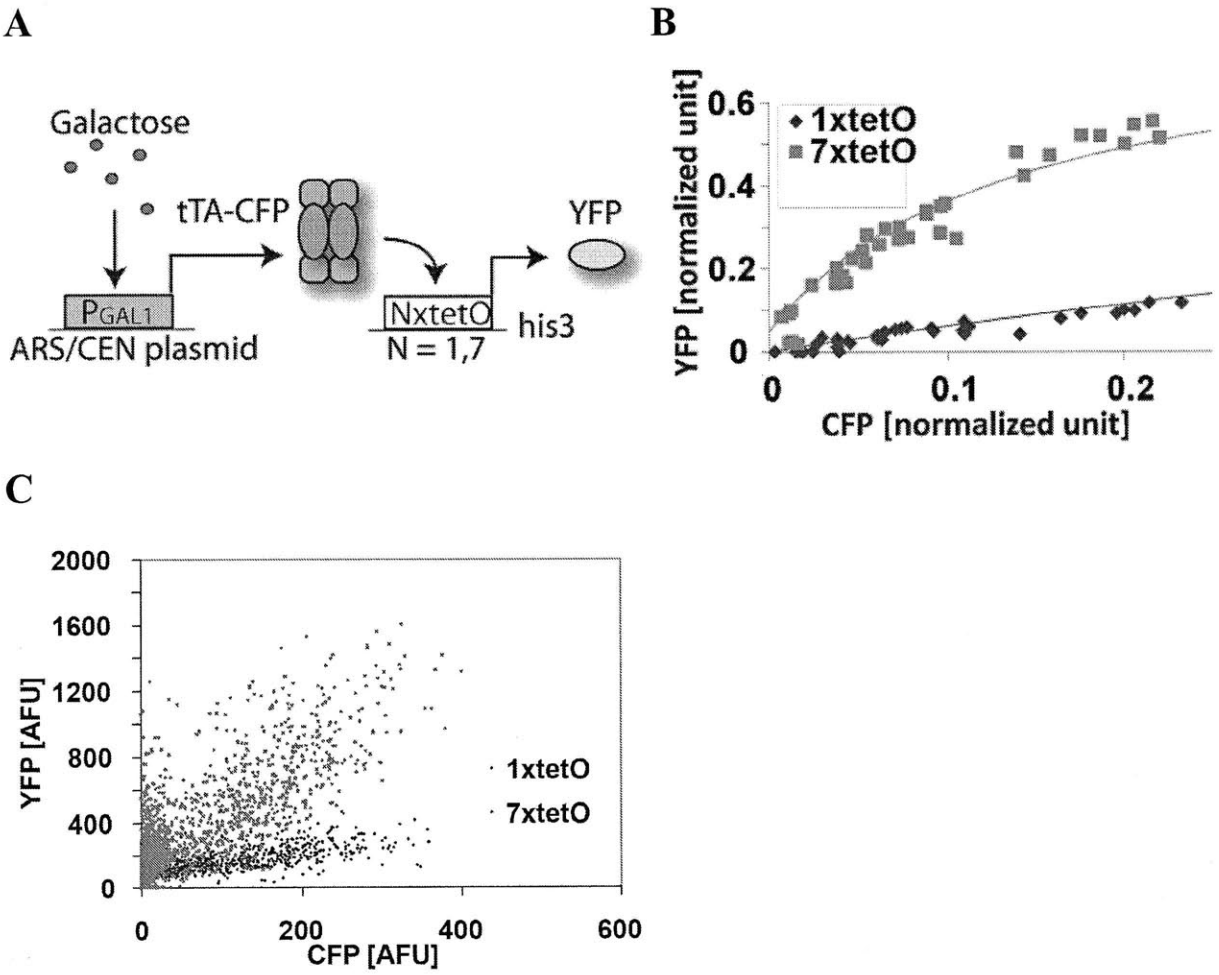
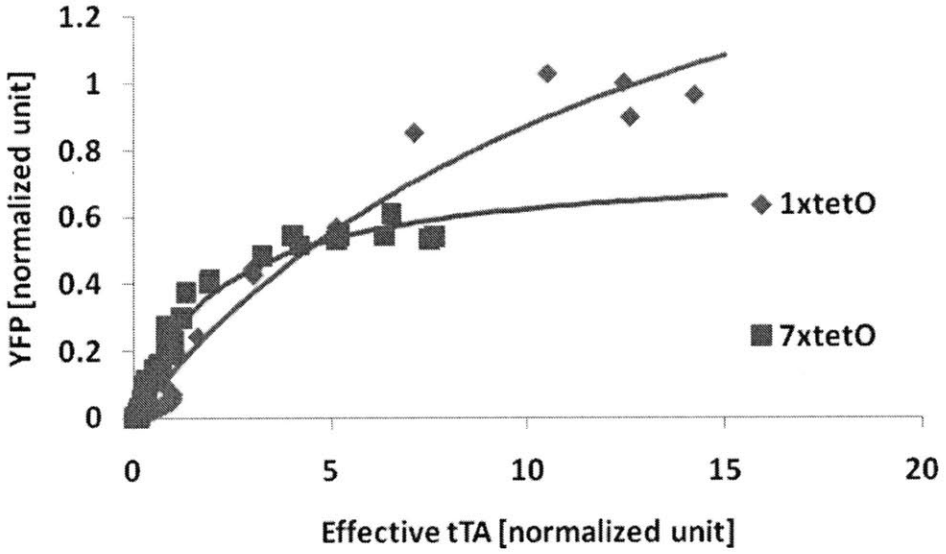


Figure 4.7. The open-loop response of the 1xtetO and 7xtetO promoters by direct titrations. The dose response curves determined by direct titration of tTA with 0 – 0.5% galactose are non-cooperative with a Hill coefficient of ~1. (A) Activator levels were directly controlled by the *GAL1* promoter. (B) Population averaged YFP versus CFP for cultures under various galactose concentrations (0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, 0.1, 0.2, 0.3, and 0.5 %). Biological triplicates were cultured and measured. (C) YFP versus CFP at the single-cell level measured by microscopy. Individual cells were pooled from all cultures in (B).

A



B

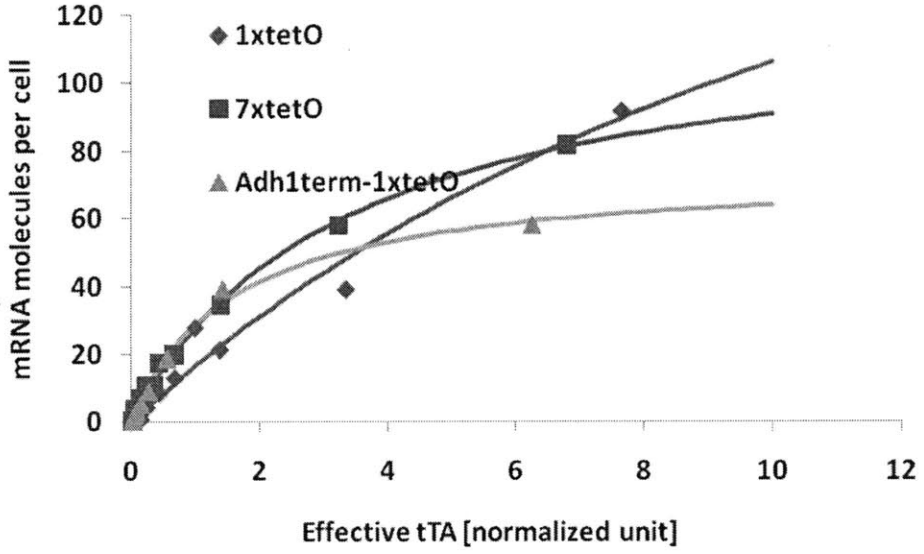


Figure 4.8. The full open-loop response across non-feedback and feedback contexts is graded with a Hill coefficient of 1. The data in both (A) and (B) were collected from multiple experiments performed on different days. The open-loop promoter response as determined by (A) the mean expression level of YFP reporter measured by fluorimetry,

FACS, or microscopy and (B) the mean number of mRNA per cell measured by FISH. One effective tTA unit is defined as the tTA level expressed from an *ADH1* promoter in the absence of doxycycline. We increased the tTA level beyond 1 effective tTA unit by expressing tTA with positive feedback. The relative tTA levels between open-loop (*ADH1* promoter) and open-loop in feedback context (1xtetO driving tTA expression) were assessed using a tTA-CFP fusion protein. The YFP measurements were normalized by the fluorescence from a positive control strain carrying integrated copies of *ADH1* promoter-YFP at the *leu2* locus.

4.4.2 The bimodal response requires a bursty promoter

As described in Section 4.2.3, a bimodal steady-state expression distribution can occur with positive feedback even when the open-loop response is graded with a Hill coefficient <= 1. This occurs when the maximum burst frequency is low, the burst size is large enough to turn the promoter on in positive feedback, and the activator regulates burst frequency (exact conditions in Figure 4.2).

Based on the model, we hypothesized that the 7xtetO promoter had a lower maximum burst frequency and larger burst size versus the 1xtetO promoter. Independent support for this hypothesis comes from a study of stochastic gene expression using the tet-OFF system in mammalian cells (Raj *et al.*, 2006). As in that study, we combined mRNA FISH measurements (see Section 3.2.2 for details) with a stochastic model of gene expression to determine the burst statistics of mRNA (see Sections 3.2.3 and 4.3.2 for details). The 7xtetO promoter has a burst size 2-fold higher than the 1xtetO promoter (Figure 4.11) while the mRNA distributions are consistent with burst frequency regulation (Figures 4.9 and 4.10). Both mRNA and intrinsic protein noise scale with the inverse square root of abundance (Figures 4.12 and 4.13, and Table 4.1), suggesting that

mRNA noise is dominated by intrinsic fluctuations (Elowitz *et al.*, 2002; Bar-Even *et al.*, 2006) and tTA regulates burst frequency. Although there is sizeable extrinsic noise in protein abundance as measured by fluorescence proteins, yet it does not appear to have a major effect (see Section 4.5).

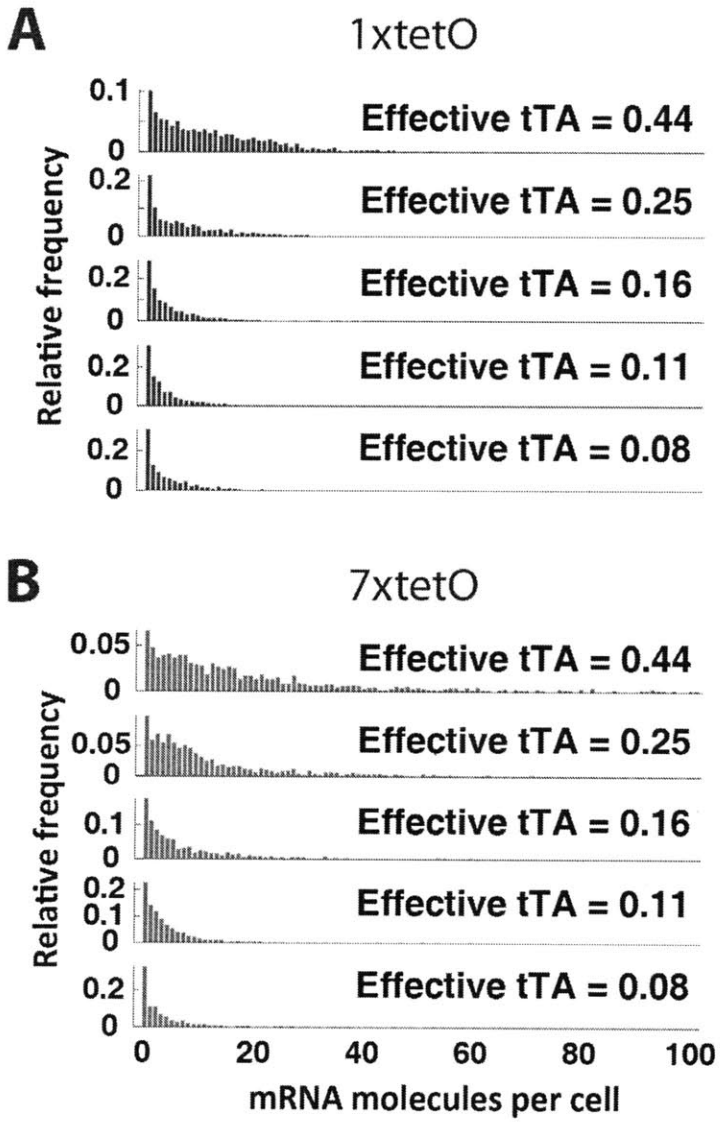


Figure 4.9. Steady-state mRNA distributions of the 1xtetO and 7xtetO promoters determined by mRNA FISH. Shown are the steady-state mRNA distributions under

250, 500, 750, 1000 and 1250 ng/mL of doxycycline. The effective tTA level was calculated based on the doxycycline concentration using equation [4.2]. These distributions were analyzed in using a stochastic model of gene expression [3.1] to generate the bursting statistics in Figures 4.10 and 4.11 and also used to calculate the mRNA noise in Figure 4.12.

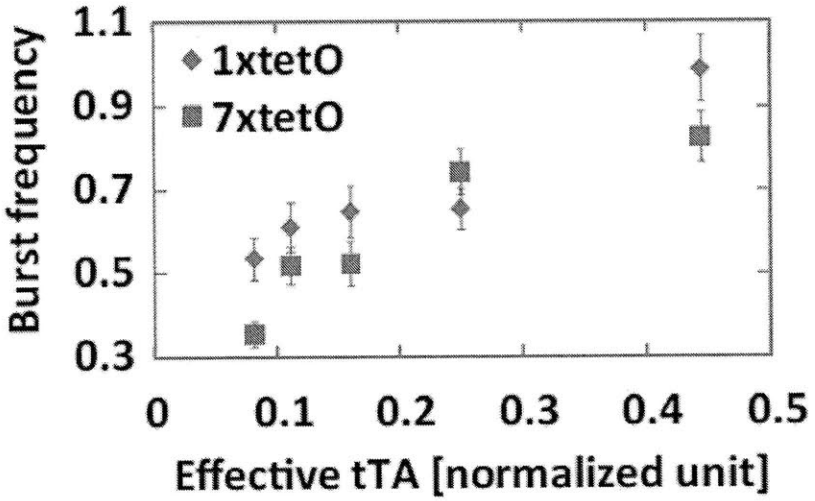


Figure 4.10. The mRNA burst frequency of the 1xtetO and 7xtetO promoters. Error bars reflect 95% CI's for maximum likelihood estimates.

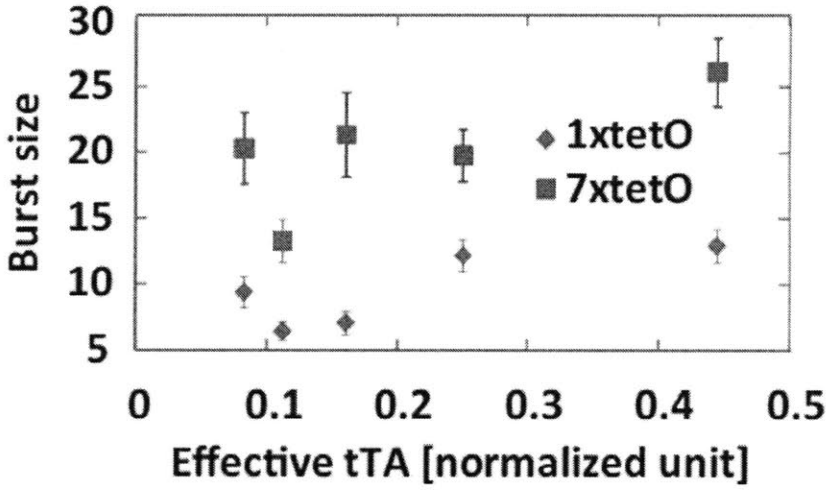


Figure 4.11. The mRNA burst size of the 1xtetO and 7xtetO promoters. Error bars reflect 95% CI's for maximum likelihood estimates.

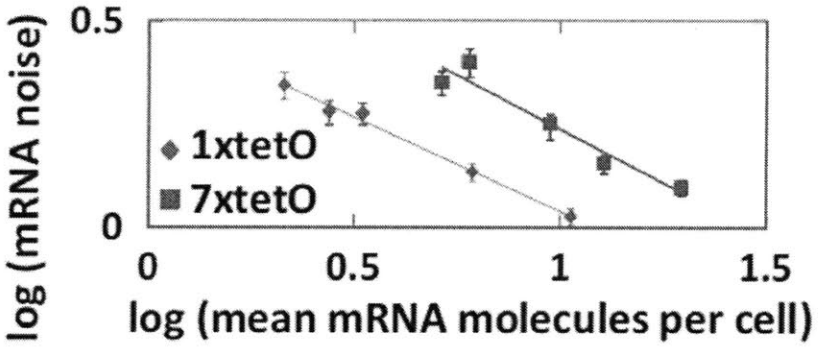
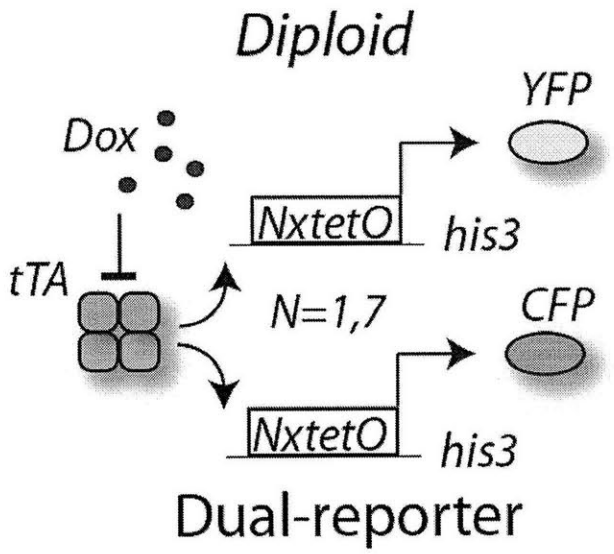
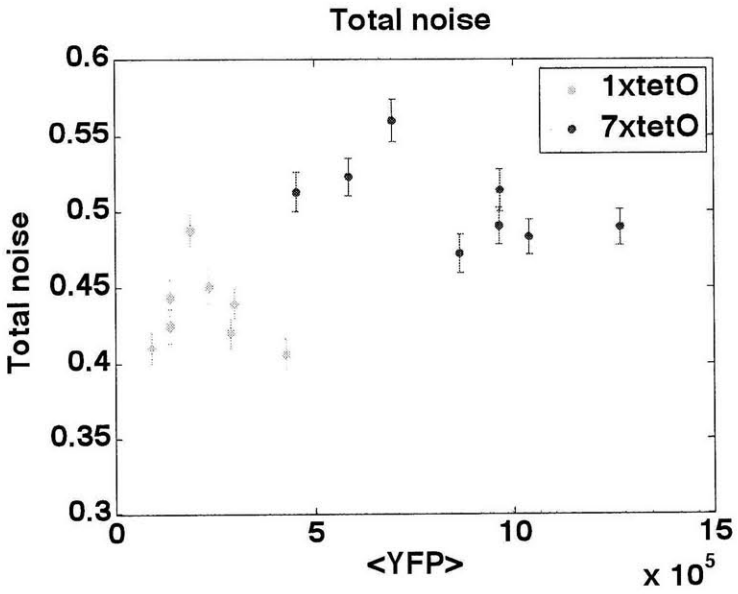


Figure 4.12. The scaling law between mRNA noise and mean mRNA level. A slope of ~-0.5 on the log-log plot suggests mRNA noise is primarily due to intrinsic sources and burst frequency regulation. Error bars representing 95% CI's were obtained by bootstrapping.

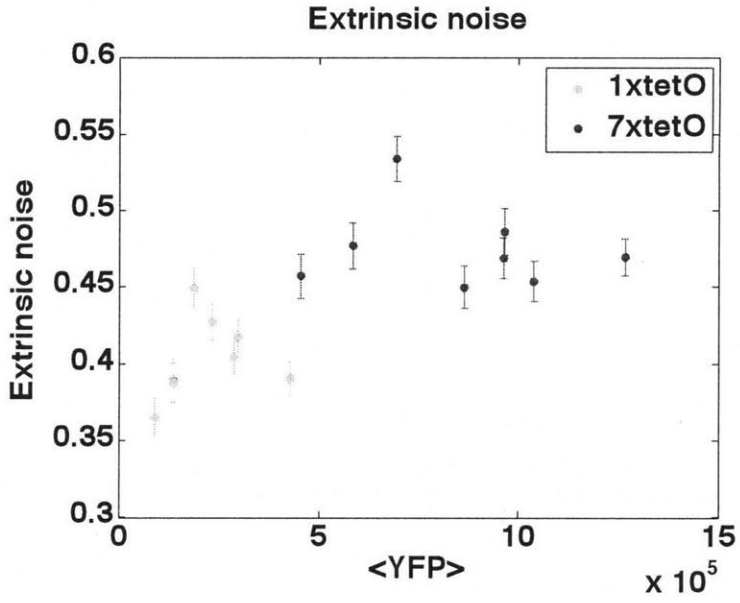
A


Diploid
YFP
Dox
NxtetO
his3
tTA
N=1,7
CFP
NxtetO
his3
Dual-reporter

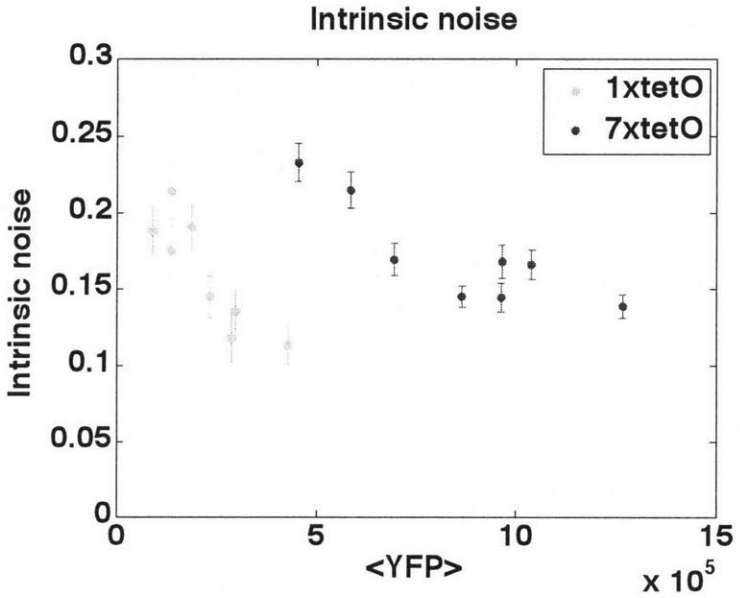
B


Total noise
1xtetO
7xtetO
Total noise
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0
5
10
15
<YFP>
x 10^5

C


Extrinsic noise
1xtetO
7xtetO
Extrinsic noise
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0
5
10
15
<YFP>
x 10^5

D


Intrinsic noise
1xtetO
7xtetO
Intrinsic noise
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
<YFP>
x 10^5

E

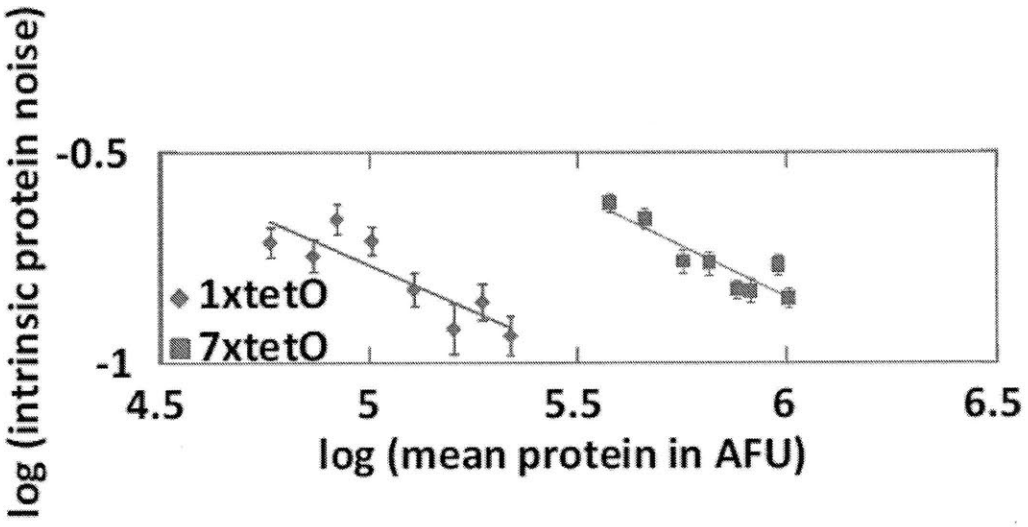


Figure 4.13. Protein noise measurement in diploid strains using the dual-reporter assay. YFP and CFP were expressed from identical tetO promoters at the *his3* locus on homologous chromosomes as diagramed in (A). Fluorescence levels were measured by microscopy and the intrinsic and extrinsic noise components were calculated as in (Swain *et al.*, 2003). Total (B), extrinsic (C) and intrinsic (D) noise as functions of population averaged expression. The mean expression was modulated by titrating tTA with 0 – 500 ng/mL doxycycline. (E) The scaling law (slope ~ -0.5 on the log-log plot) between protein noise and mean protein level is consistent with burst frequency regulation. Error bars representing 95% CI's were obtained by bootstrapping.

Table 4.1. Comparison of regression coefficients

Open-loop mRNA noise (Figure 4.12)	1xtetO	7xtetO
Scaling exponent	-0.45±0.05	-0.56±0.14
Relative burst size	1.00±0.34	2.00±0.44

Open-loop intrinsic protein noise (Figure 4.13-E)	1xtetO	7xtetO
Scaling exponent	-0.44±0.13	-0.48±0.12
Relative burst size	1.00±0.18	4.46±0.20

In the open-loop context, the *ADH1* promoter-driven tTA level was never high enough to saturate either promoter. Therefore, we used the closed-loop data, where tTA levels are much higher, to estimate the maximum burst frequency (Figure 4.8). Consistent with our hypothesis and the mammalian study (Raj *et al.* 2006), 7xtetO has a 5-fold lower maximum burst frequency than 1xtetO (Table 4.4). Taken together, multiple tetO binding sites make a promoter more sensitive to tTA, with a lower burst frequency and a higher burst size. We are unable to determine if the increased burst size is due to a longer duration burst or a more intense burst (Figure 4.4).

Table 4.2. Parameters from fitting the gene regulatory function

	YFP		mRNA	
	k_{max}	K	k_{max}	K
1xtetO	2.10±0.55	14.1±6.2	272±220	15.6±17.2
7xtetO	0.75±0.06	2.08±0.39	121±11	3.38±0.62

Interestingly, a 1xtetO promoter variant containing the yeast *ADH1* terminator upstream of the 1xtetO site has a non-cooperative open-loop response (Figure 4.15) and also exhibits a bimodal response in positive feedback (Figure 4.14). This promoter has a higher burst size and lower maximum burst frequency compared to 1xtetO (Figure 4.16 and Table 4.3). Therefore, as expected by the theory, the noise properties of the promoter are responsible for the bimodal response and not multiple binding sites, per se.

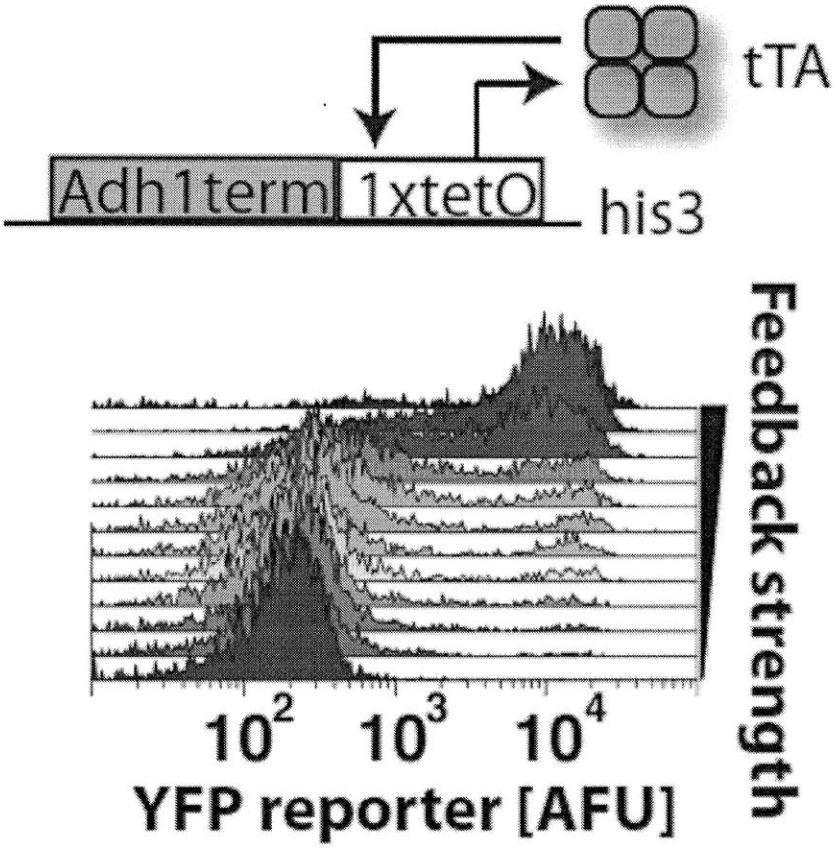
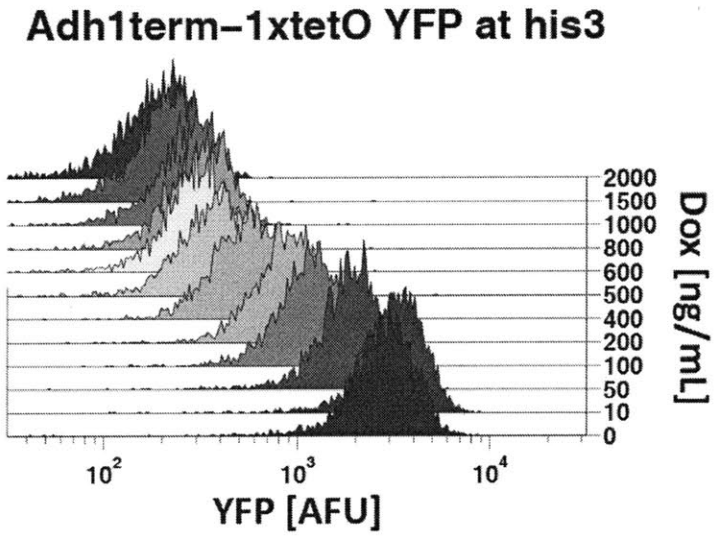
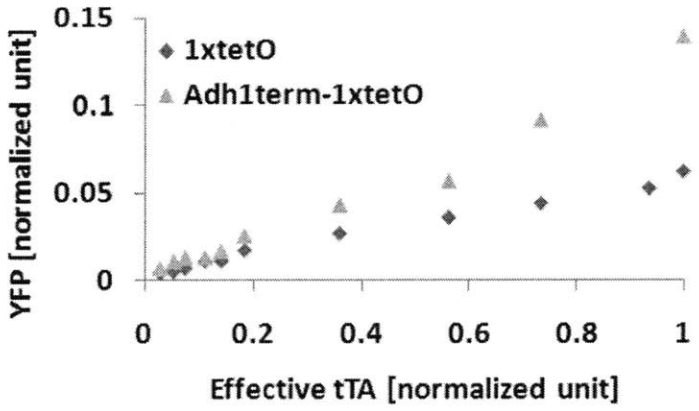


Figure 4.14. Closed-loop response of the ADH1term-1xtetO promoter is bimodal. The Adh1term-1xtetO promoter was created by inserting an *ADH1* terminator upstream of 1xtetO in the sense direction. The feedback strength was modulated by varying doxycycline from 0 to 1000 ng/mL.

A


Adh1term–1xtetO YFP at his3
2000
1500
1000
800
600
500
400
200
100
50
10
0
Dox [ng/mL]
10^2
10^3
10^4
YFP [AFU]

B


0.15
0.1
0.05
0
1xtetO
Adh1term-1xtetO
YFP [normalized unit]
0
0.2
0.4
0.6
0.8
1
Effective tTA [normalized unit]

C

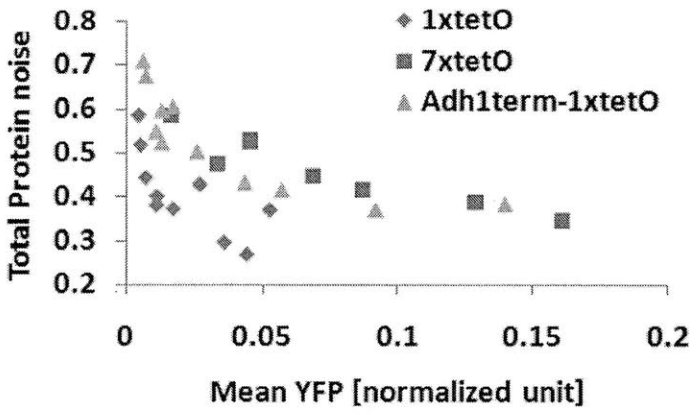
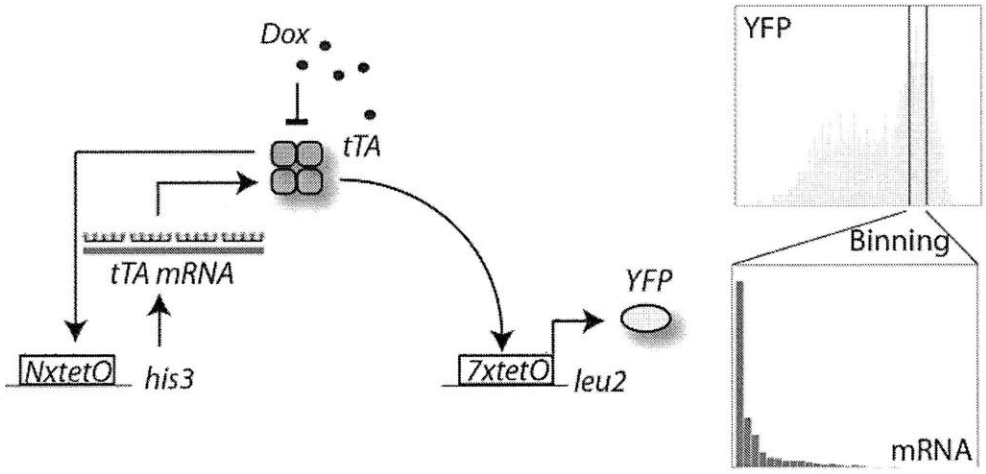


Figure 4.15. Comparison of open-loop response between 1xtetO and Adh1term-1xtetO promoters. An *ADH1* promoter was used to express tTA constitutively from a yeast centromeric plasmid and the effective tTA level was calculated from doxycycline concentration using equation [4.2]. The Adh1term-1xtetO promoter has (A) graded steady-state response, (B) a Hill coefficient of ~1 (see also Figure 4.8 for full response) and (C) higher total protein noise.

A



B

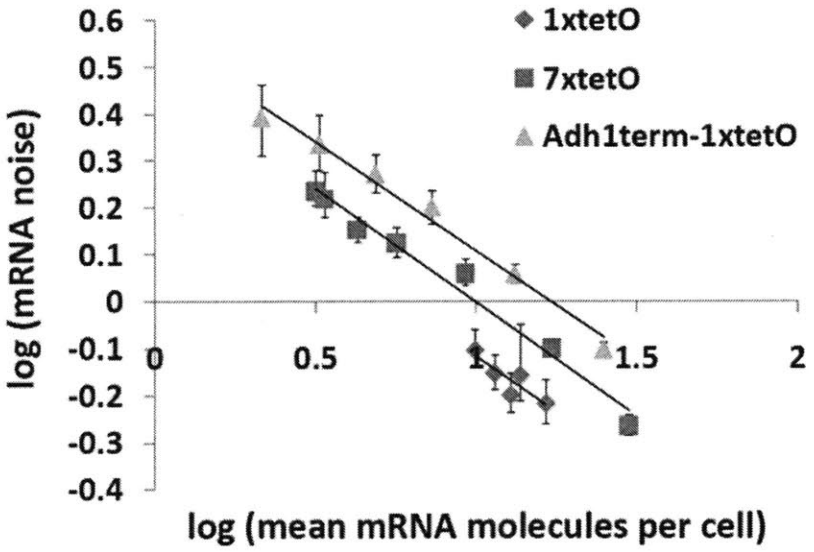


Figure 4.16. Measuring mRNA noise in closed-loop as a function of TF level. The burst statistics of mRNA in closed-loop configuration was determined by measuring the mRNA distributions of tTA within a bin of YFP expression (A). A tight bin of YFP represents a relatively constant tTA level across one cell-cycle (B). The expected scaling law (a slope ~ -0.5 on the log-log plot) for mRNA is observed in closed-loop configuration, implying that burst frequency is being modulated by the transcriptional activator. As in the open-loop case 7xtetO and Adh1term-1xtetO are noisier than 1xtetO. Error bars were obtained by bootstrapping.

Table 4.3. Regression coefficients of closed-loop mRNA noise (Figure 4.16)

	1xtetO	7xtetO	Adh1term 1xtetO
Scaling exponent	-0.47±0.22	-0.48±0.04	-0.46±0.03
Relative noise strength	1.00±0.80	1.35±0.49	1.67±0.47

4.4.3 The bimodal response requires an unstable TF

A strong prediction of our model is that stabilization of tTA will increase the maximum burst frequency and eliminate the bimodal expression pattern. Global measurements of protein stability in yeast reveal that TFs tend to be less stable than typical proteins (Belle *et al.*, 2006). While tTA (a fusion between tet repressor and the VP16 activation domain) stability has not been determined, the *in vivo* half-life of a lexA-VP16 fusion protein in yeast was 6 minutes (Salghetti *et al.*, 2001). Moreover, mono-ubiquitination of lexA-VP16 was required to license the activator for transcriptional initiation, but lead to subsequent polyubiquitination via Met30p, targeting it for degradation (Salghetti *et al.*, 2001). To stabilize tTA, we expressed the active mono-ubiquitinated version and deleted *MET30*. *MET4* must also be deleted, as single delection of *MET30* is lethal (Salghetti *et al.*, 2001). As expected, the closed-loop response of the stabilized tTA is more graded (Figure 4.17).

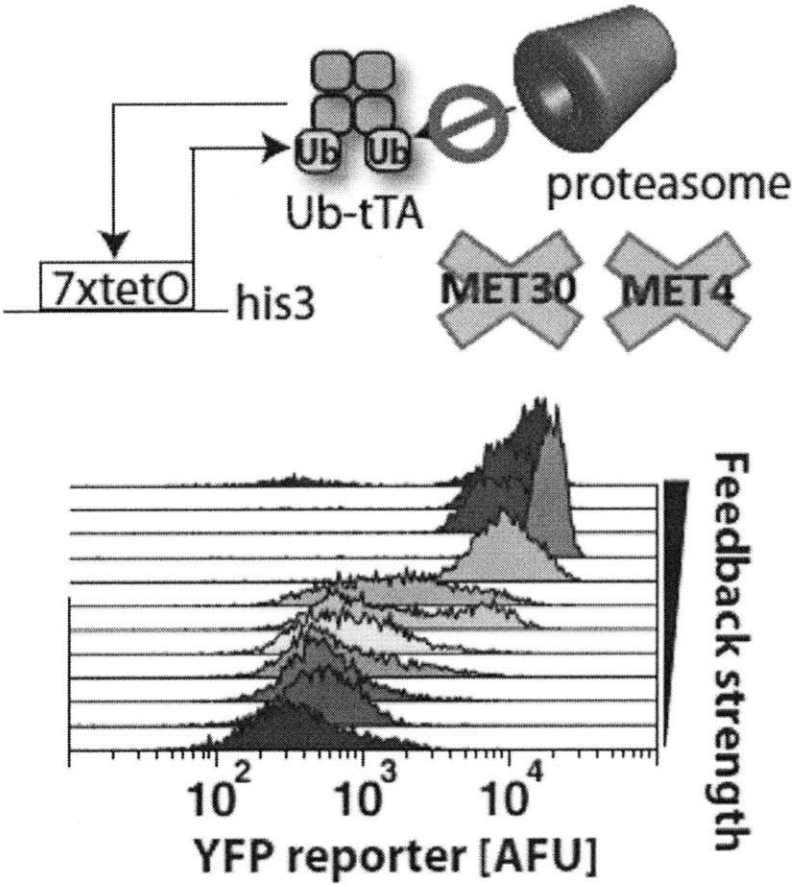


Figure 4.17. Stabilizing transcription factor reduces stochastic fluctuation and eliminates bimodality. To create a stabilized TF, an ubiquitin-tTA fusion was expressed in a *met30Δ met4Δ* background (upper panel). The closed-loop reporter response measured by flow cytometry becomes more graded for 7xtetO (lower panel). Feedback strength was modulated by varying doxycycline as in Figure 4.5.

To verify whether a stochastic model could quantitatively describe our results, we measured tTA mRNA and protein half-lives. To determine the mRNA stability, we stopped transcription in a 7xtetO closed-loop strain by adding doxycycline and the transcriptional inhibitor thiolutin. Cells were fixed at specific time points after inhibition and tTA and YFP mRNA abundance was measured by FISH. Both transcripts have a

half-life ~15-20 minutes (Figure 4.18). We determined the stability of tTA at the population level with promoter shut-off experiments followed by Western blotting (Figure 4.19). Surprisingly, tTA appears to have a ~70 minute half-life, much longer than the 6 minute half-life of lexA-VP16 (Salghetti *et al.*, 2001). With this half-life the maximum burst frequency is too high and the Friedman model cannot even qualitatively describe bimodal expression.

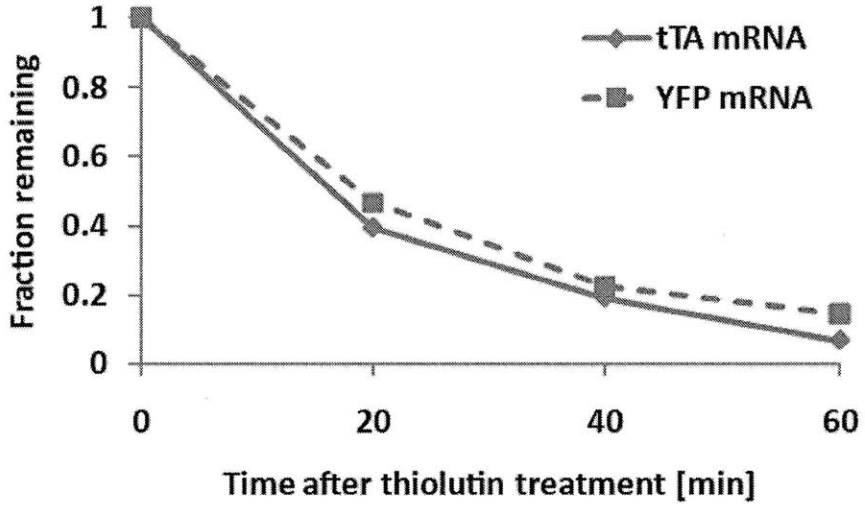


Figure 4.18. Stability of mRNA. We grew a 7xtetO closed-loop strain (Y161) to steady state and stopped transcription by addition of high levels of doxycycline and the transcriptional inhibitor thiolutin. Cells were fixed at specific time points after inhibition and the abundance of both tTA and YFP mRNAs were measured by FISH. Lines are for the eye. We found that both transcripts follow first-order decay kinetics and have a ~15-20 minute half-life (Table 4.4).

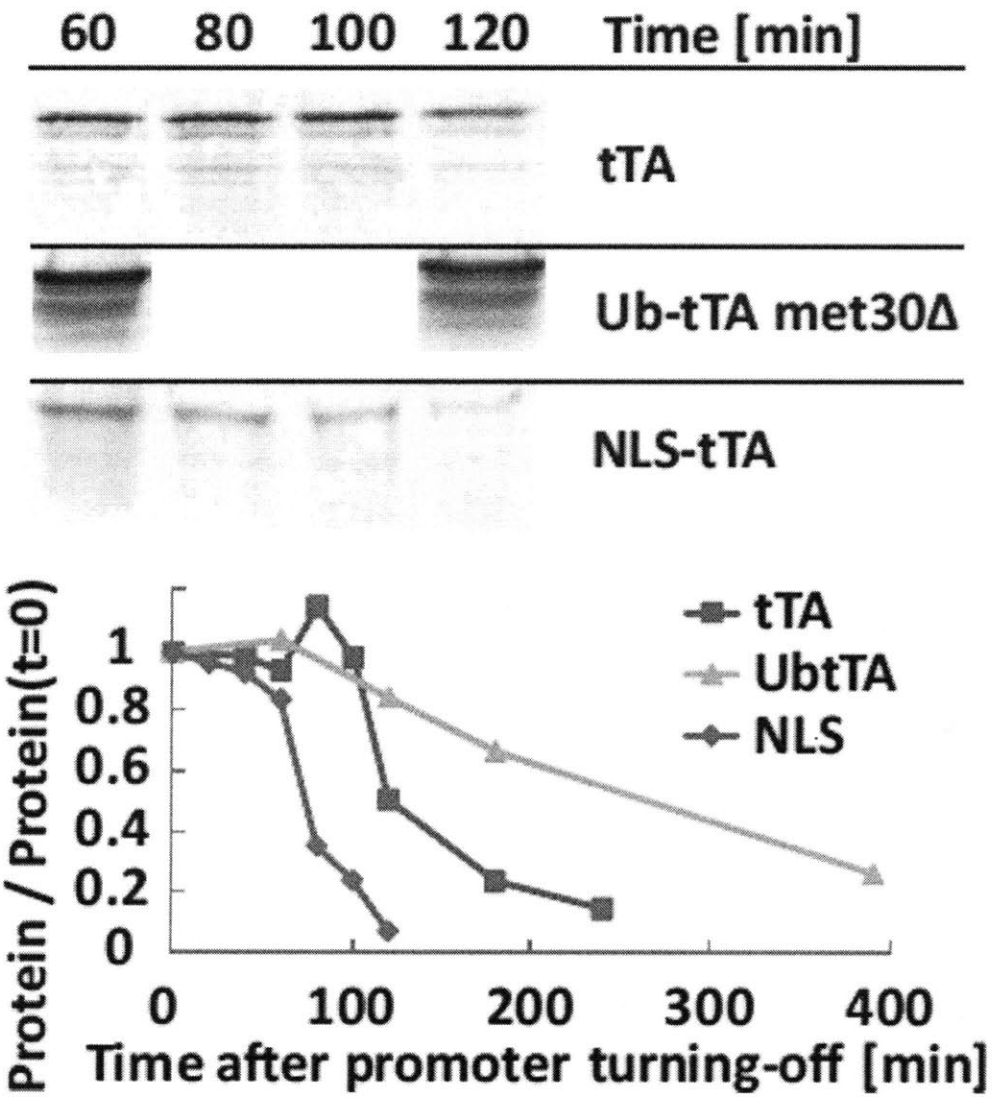


Figure 4.19. Stability of tTA variants. To determine tTA stability, its expression was halted via a galactose–inducible promoter and levels were followed by quantitative western blotting. Lanes contain equivalent amount of total protein as determined by Bradford Assay. After accounting for the 60-minute lag required for mRNA depletion and tTA folding, tTA was found to have a ~70 minute half-life. The stabilized tTA variant has a ~175 minute half-life similar to the doubling time. The SV40 NLS-tagged tTA has a ~15 minute half-life, suggesting that nuclear transport was rate limiting.

4.4.4 The importance of nuclear transport

Given the short half-life of the lexA-VP16 fusion, we suspected that multiple forms of tTA were present in the cell and only the active form was unstable. Because activation-coupled degradation of tTA occurs only in the nucleus, we hypothesized that cytoplasmic tTA was stable and nuclear transport of tTA was slow. Nuclear transport of the reverse (r)tTA has been shown to be limiting in yeast, and addition of nuclear localization signals (NLS) alters its subcellular distribution (Becskei *et al.*, 2004).

To test the effect of nuclear transport, we fused a mammalian NLS to tTA, confirmed the nuclear localization (Figure 4.20), and measured the open- and closed-loop responses. As before, the NLS-tTA open-loop response was graded and non-cooperative (Figure 4.21). The closed-loop responses remained bimodal for 7xtetO (Figure 4.22) although more cells were found with intermediate expression levels compared to tTA (Figure 4.5). NLS-tTA stability was reduced to a ~10 minute half-life (Figure 4.19), suggesting that the increased stability of tTA was indeed due to a stable cytoplasmic fraction not susceptible to rapid degradation. Incorporating both nuclear transport and the shorter tTA half-life in our model (Figure 4.23), we can recapitulate the closed-loop responses for NLS-tTA (Figure 4.24) using measured *in vivo* parameters for the open-loop promoters (Table 4.4). The more pronounced bimodal response from nuclear transport-limited tTA (Figure 4.5) compared to NLS-tTA remains unclear, but likely depends on details of tTA transport. Finally, our preliminary simulations with this type of cell-cycle dependent nuclear transport of tTA indicate an increased separation of ON and OFF peaks (Figure 4.25) but not near what we measured experimentally.

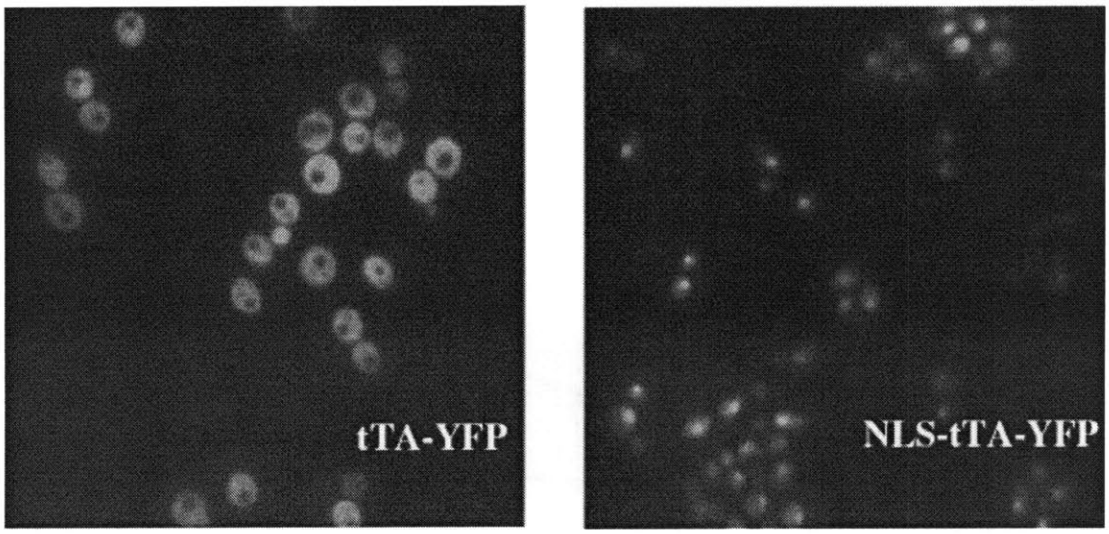
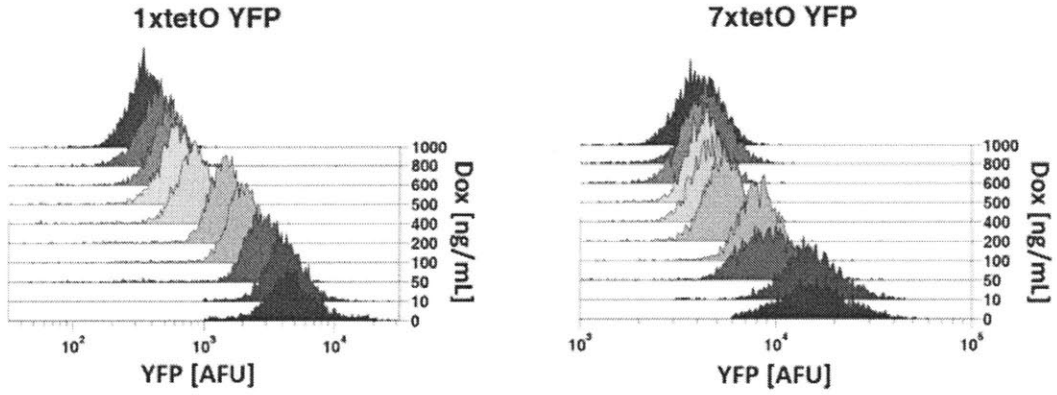


Figure 4.20. Localization of untagged and SV40-NLS tagged tTA-YFP fusion proteins. Both tTA variants were constitutively expressed from an *ADH1* promoter in the absence of doxycycline. The black "holes" in the tTA-YFP panel are vacuoles, NOT the nucleus.

A



B

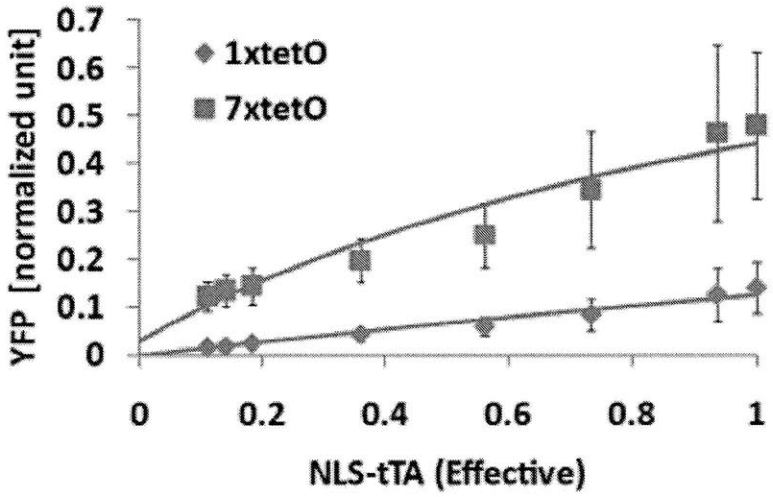


Figure 4.21. Open-loop response of NLS-tTA is graded. An *ADH1* promoter was used to express NLS (SV40)-tTA constitutively from a yeast centromeric plasmid and doxycycline was added to reduce the active fraction. (A) The steady-state expression distributions of both 1xtetO and 7xtetO promoters are graded and unimodal. (B) The promoter response curves are well fit to a Hill coefficient of ~1 and resemble those of the regular tTA (Figure 4.6). Error bars indicate s.d. of triplicate samples.

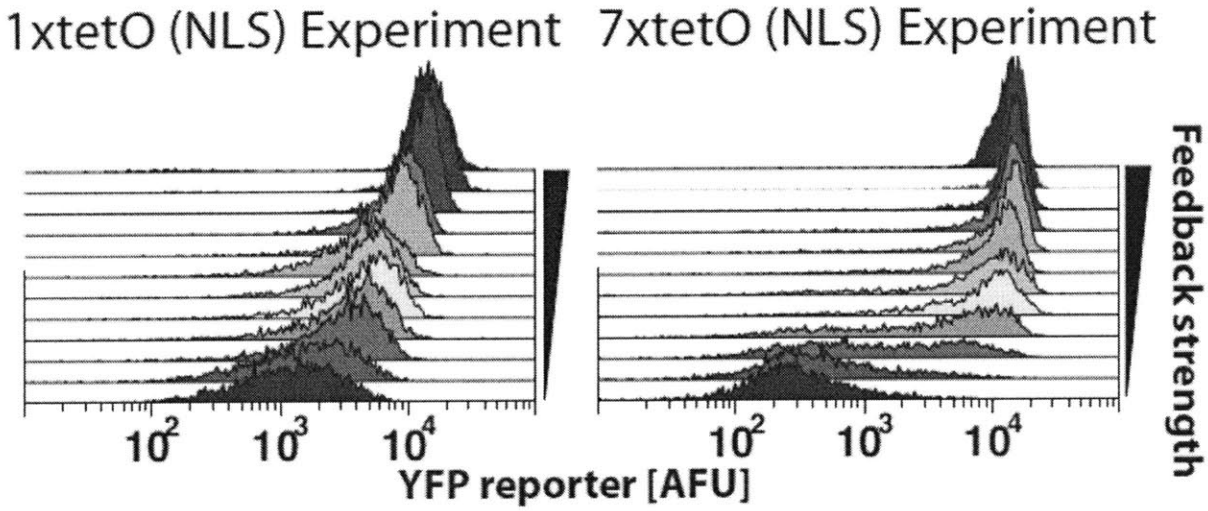


Figure 4.22. Closed-loop response of NLS-tTA. A SV40 NLS was fused to the N-terminal of tTA and the closed-loop reporter response at steady-state was measured by flow cytometry. The feedback strength was modulated by varying doxycycline as in Figure 4.5. The measured expression profiles for NLS-tTA remain bimodal, but less so compared to normal tTA.

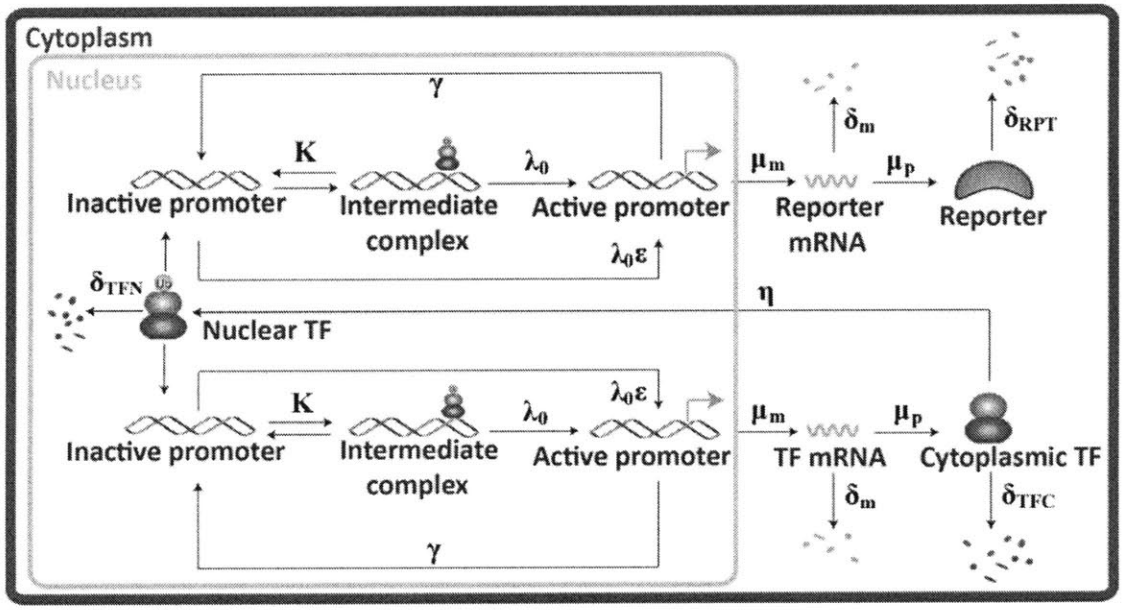


Figure 4.23. The schematic of the full stochastic model. All the states and elementary steps considered are shown. The description of each parameter can be found in Table

4.4. This model was used to simulate the steady-state reporter distributions in Figures 4.24 and 4.25.

Table 4.4. Parameters used in the stochastic model

	Description	Value / Error		Unit
		1xtetO	7xtetO	
λ_o	Maximum TF mediated promoter activation	25 (16, 39)	5 (3.5, 6.8)	δ_m
ε	Basal burst factor	0.04±0.02	0.015±0.01	–
γ	Promoter inactivation	50 (40, 74)	25 (20, 32)	δ_m
K	TF-promoter dissociation constant	12500±5500	2500±460	# Nuclear TF
μ_m	Transcription	500 [a]		δ_m
μ_p	Translation	23.1 [b]		min^{-1}
η	Nuclear translocation	0.2 [c]		min^{-1}
δ_m	mRNA degradation	0.04±0.009		min^{-1}
δ_{RPT}	Reporter degradation	0.007 [d]		min^{-1}
δ_{TFC}	Cytoplasmic TF degradation	0.007 [d]		min^{-1}
δ_{TFN}	Nuclear TF degradation	0.07±0.03[e]		min^{-1}
K	Range of dissociation constants	$K_{no\ dox}$ to $15K_{no\ dox}$ [f]		# Nuclear TF

a. γ and μ cannot be estimated independently

b. Average translation rate in *S. cerevisiae* based on (Ghaemmaghami *et al.*, 2003; Bell *et al.*, 2006)

c. Inferred from the nuclear-cytoplasmic partitioning of NLS-tTA and (Becskei *et al.*, 2004)

d. Assumed equal to the dilution rate

e. Obtained by fitting the Western blot data in Fig 3C to a biexponential rate describing the process: Cytoplasmic TF → Nuclear TF → Ø. A nuclear translocation rate of 0.2 min^{-1} was assumed.

f. Inferred from equation [4.2] and range of doxycycline concentrations used in our experiments

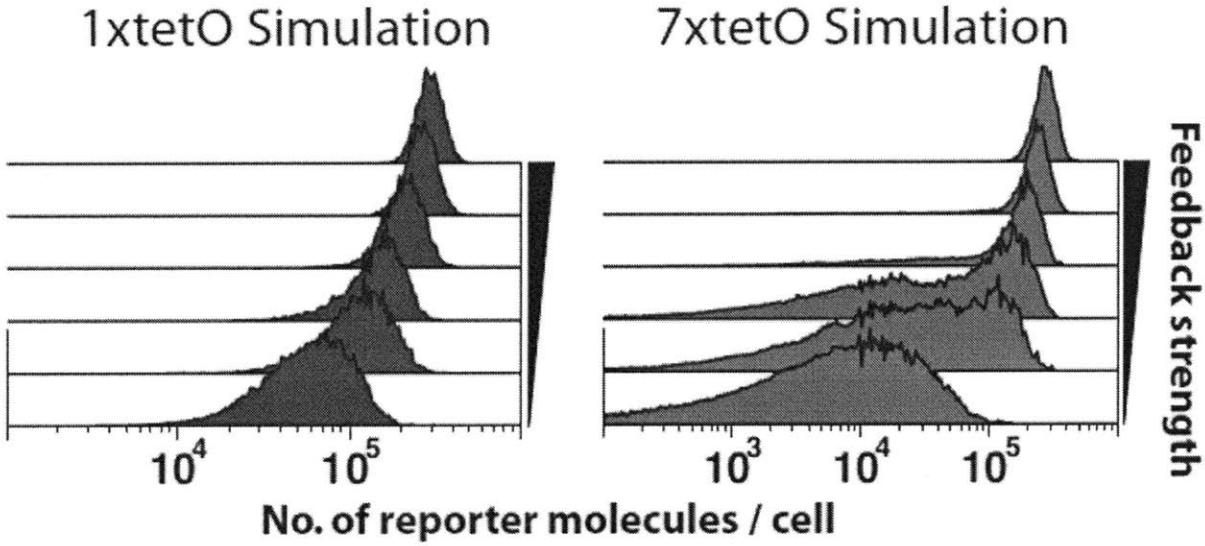


Figure 4.24. Simulation of closed-loop response. Stochastic simulations were performed using the scheme in Figure 4.23 and parameter values in Table 4.4. Results capture differences between the graded versus bimodal expression profiles between 1xtetO and 7xtetO and generally agree with experimental closed-loop NLS-tTA data (Figure 4.22).

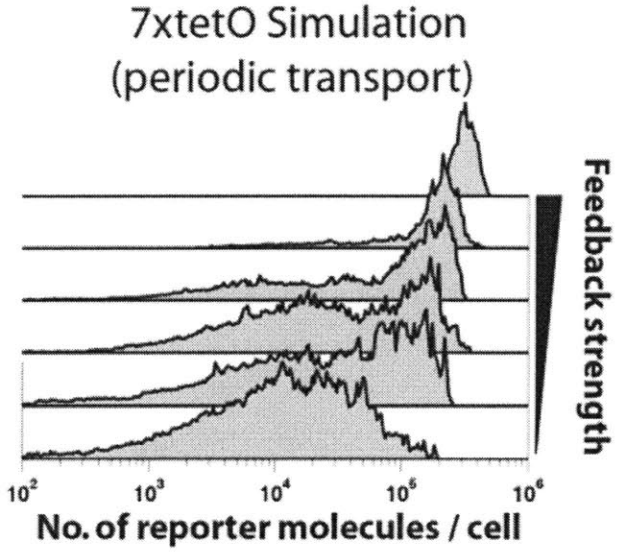


Figure 4.25. Simulation of closed-loop response with a periodic, cell-cycle dependent nuclear transport rate. Period nuclear transport leads to a better separation between the ON and OFF states of reporter expression. The periodic nuclear translocation rate was modeled as $\eta(t) = 0.1\cos(1+2\pi t/100)\,\mathrm{min}^{-1}$. The 7xtetO parameter set in Table 4.4 was used.

4.5 Discussion

Surprisingly, a stochastic model that only accounted for intrinsic noise was able to describe our experimental results when the NLS-tTA was employed. Extrinsic noise is significant in fluorescent reporter proteins, but the corresponding mRNA noise appears completely intrinsic in nature. This suggests the source of extrinsic noise is likely due to global fluctuations in translational expression capacity (Colman-Lerner *et al.*, 2005). Clearly, in comparing 1xtetO to 7xtetO, intrinsic noise is necessary for bimodal expression, but because protein feeds back on its own expression, extrinsic noise might be expected to play a role in dictating the steady-state response. Introducing slow, uncorrelated fluctuations in translational capacity with an autocorrelation time of the cell cycle cause should cause both the ON and OFF peaks to widen, leading to less well separated distributions. It may be that extrinsic noise plays less of a role because the variation in fluorescent protein we measured does not accurately reflect variation in nuclear tTA activity. It is also formally possible that extrinsic noise in translational capacity is somehow correlated with whether the feedback loop is ON or OFF, although we have no evidence in support of this idea.

In principle, the result in Figure 4.17 should remove any lingering doubts about the non-cooperative open-loop response of 7xtetO. If the response were truly ultrasensitive and the measurements in Figures 4.6 and 4.7 were incorrect, than stabilizing the activator would not abolish the bimodality but rather alter the range of positive feedback strength over which it occurs.

Another possibility is that non-linear degradation or dilution of the TF results in a deterministic bistability. A recent study (Tan *et al.*, 2009) demonstrated that in a synthetic positive feedback circuit of T7 RNA polymerase in *E. coli,* bistable gene expression can be generated with non-cooperative activation because of non-linear growth retardation caused by high levels of T7 RNA polymerase. We do not consider this mechanism as the cause for bimodal expression in our system for multiple reasons. First, no growth retardation was observed with the 7xtetO positive feedback strain in the bimodal regime (500-1000 ng/mL doxycycline). Second, although both 1xtetO and 7xtetO positive feedback strains exhibited slight growth retardation (a ~120-150 minute doubling time versus ~100 minute in the normal case) when cultured with 0-200 ng/mL doxycycline, no bimodality was observed with 1xtetO positive feedback. In addition, the 1xtetO promoter variant containing an *ADH1* terminator upstream showed no retarded growth at 0-200 ng/mL doxycycline, and yet it yielded bimodal expression profile in this regime. Third, the degradation rates of the TF (Figure 4.19) are first order with respect to TF levels.

We were not able to completely account for the better separations between the ON and OFF states in the case of tTA with our model. One possibility is that multiple steps or delays involved in the nuclear transport of tTA increases the stability. When mRNA and protein half-lives are comparable, a 3-stage model with explicit transcription

and translation does lead to better separation and stability versus a 2-stage model with coupled transcription/translation (Section 4.2.3). This, however, is related to the faster degradation rates of the 3-stage model compared to the "equivalent" 2-stage model. The effect of delays or additional first order steps has little effect on the distribution. A second possibility is a true steady-state has not been reached in our experiments and the bimodal expression observed is a transient phenomenon. A previous set of studies demonstrated that the HIV-1 Tat positive feedback loop is non-cooperative and can yield a bimodal transient response (Weinberger *et al.*, 2005; Weinberger and Shenk, 2007; Weinberger *et al.*, 2008). The Tat system possesses a single OFF steady-state that must be transiently excited by Tat addition and eventually relaxes. It is unlikely for our system to only have an OFF steady-state in the bimodal regime, since longer experiments (>30 hours) never resulted in a completely OFF population (Figures 4.26 and 4.27). Moreover, we did not transiently add doxcycline addition to vary the feedback strength; the concentration was kept constant.

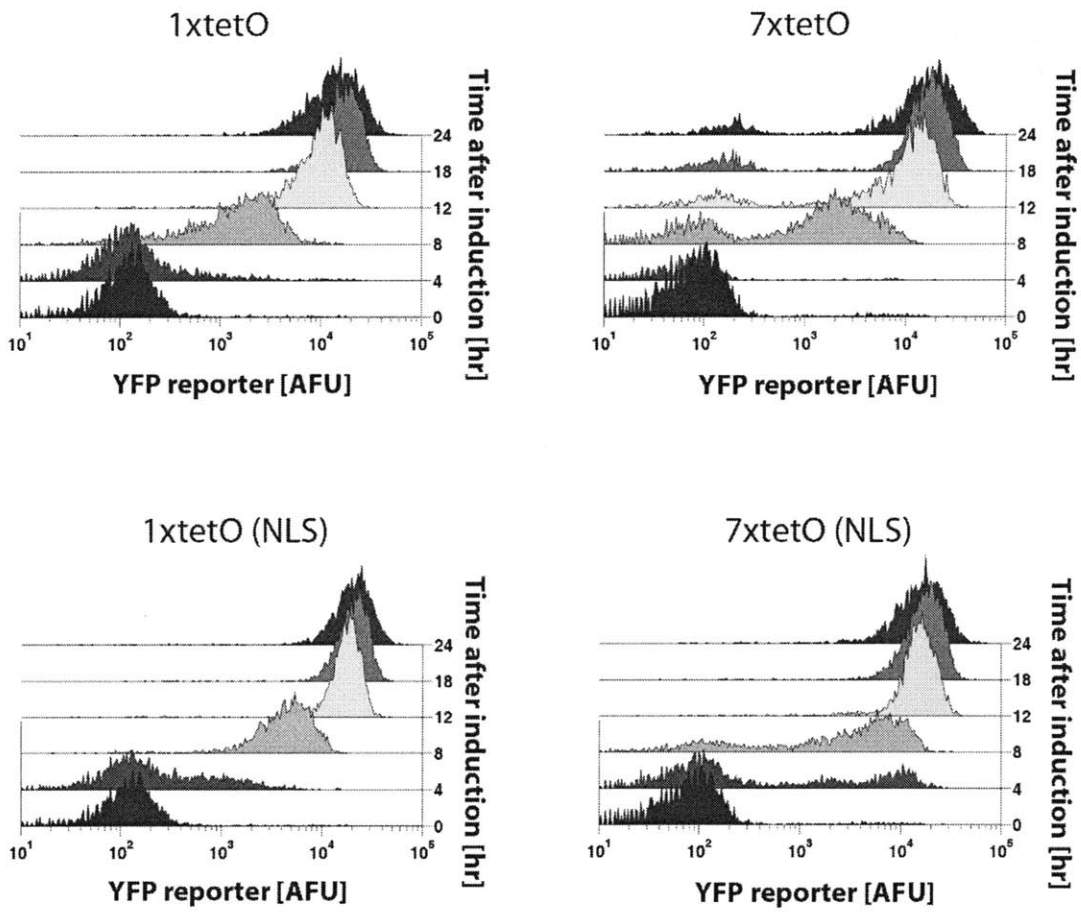
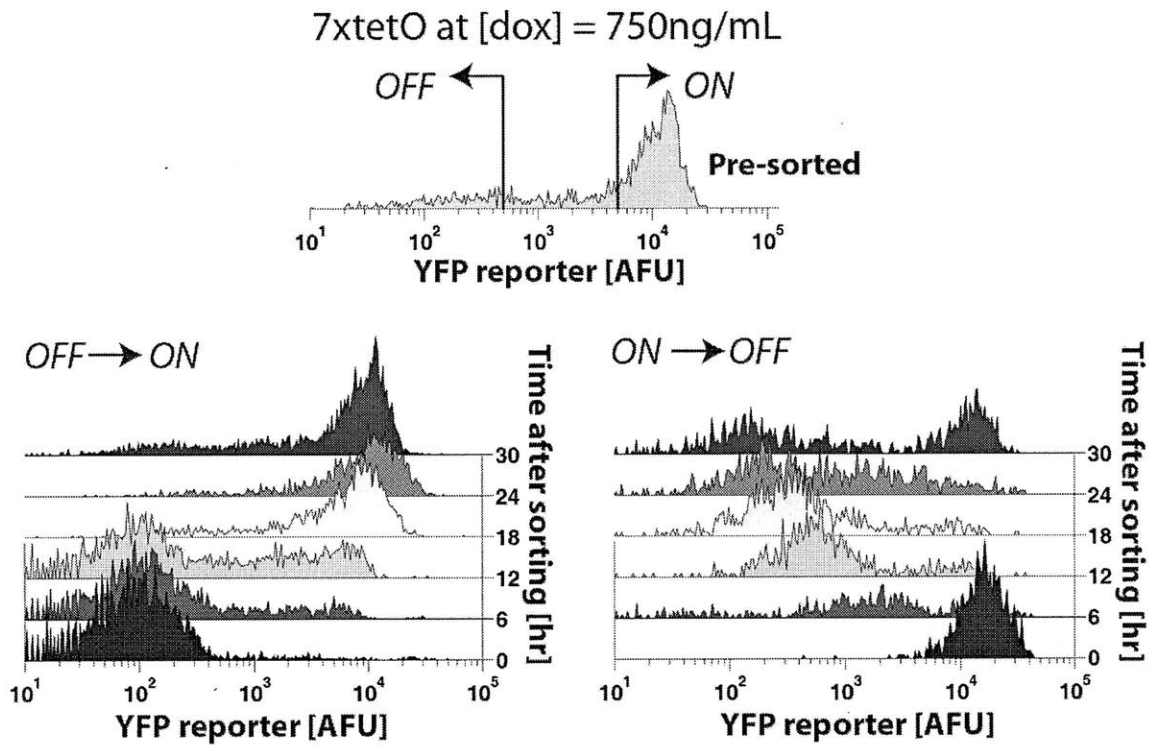


Figure 4.26. Induction kinetics in positive feedback strains. Before induction, cells were cultured in mimimal medium containing 5,000 ng/mL. To induce high tTA expression, cells were diluted in fresh medium containing 250 ng/mL doxycycline at time = 0 hour. To maintain $OD_{600} < 1.0$, cells were back-diluted at 12 hours. For both tTA and NLS-tTA, transient bimodality is observed with 7xtetO but not with 1xtetO.

A


7xtetO at [dox] = 750ng/mL
OFF
ON
Pre-sorted
YFP reporter [AFU]
OFF → ON
ON → OFF
Time after sorting [hr]
0
6
12
18
24
30

B

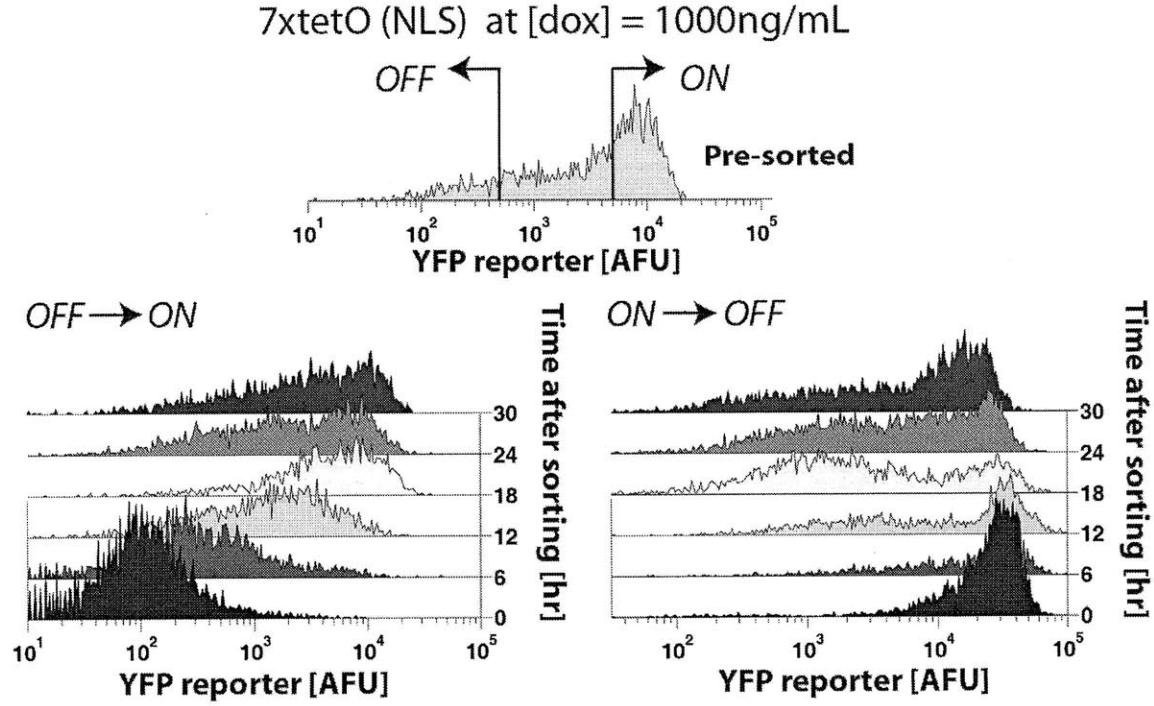

7xtetO (NLS) at [dox] = 1000ng/mL
OFF
ON
Pre-sorted
YFP reporter [AFU]
OFF → ON
ON → OFF
Time after sorting [hr]
0
6
12
18
24
30

Figure 4.27. Stochastic dynamics switching of bimodal population. (A) An initial bimodal population of 7xtetO positive feedback strain grown at 750 ng/mL doxycycline was sorted by fluorescent activated cell sorting (FACS) into ON and OFF populations defined by low and high levels of YFP. Cells that were undergoing transition between the two states, as indicated by an intermediate expression level, were not considered. For the initially OFF population, steady-state distribution was recovered between 24-30 hours. For the initially ON population, relaxation to the steady-state required more than 30 hours. (B) The same experiment was performed with NLS-tTA bearing 7xtetO positive feedback strain grown at 1000 ng/mL doxycycline. Both populations relaxed to the steady-state within 24 hours.

A third possibility for the increased stability is variability and periodicity in the nuclear transport of tTA. While we have not directly examined these dynamics, static snapshots of cells constitutively expressing a tTA-YFP fusion indicate variability in its nuclear localization (Figure 4.20). One possibility for this variability is cell-cycle dependent transport. Fungi have a closed mitosis where the nuclear envelope does not break down, but some fungal proteins do exhibit cell-cycle nuclear transport due to enhanced nuclear permeability during mitosis (Ovechkina *et al.*, 2003). In addition, proteins in budding yeast with cyclin-dependent NLS's display cell-cycle dependent nucleocytoplasmic shuttling (Kosugi *et al.*, 2009). Whatever the effect, varying the strength of the NLS of tTA and possibly other TFs seems to have unexpected consequences on target genes when feedback loops are present.

To conclude, this work is related to examples where stochastic views of positive feedback in enzymatic cycles (Smoilov *et al.*, 2005) and spatial organization (Altschuler *et al.*, 2008) possess bimodal activity when a deterministic description admits an intermediate activity. In addition, a stochastic view of a non-cooperative positive

feedback loop involved in HIV escape from latency leads to a transient bimodal response (Weinberger and Shenk, 2007), but our work demonstrates a steady-state bimodal response. The hallmarks of noise-induced bimodality in gene expression – positive feedback loops and unstable proteins – are characteristic of many TFs and promoters and likely widespread in biological systems (Table 4.5). This work also suggests that multiple binding sites may be associated with all-or-none response not by virtue of cooperative binding but because of increased noise. Finally, this work provides new guidelines for the construction of a bistable switch based on positive feedback for applications in synthetic biology and metabolic engineering.

Table 4.5. Transcription factors known to display the hallmarks of noise-induced bimodality.

Organism	TF	Evidence for direct positive autoregulation	TF half-life (minutes)	# of binding sites	Evidence for variability
B. subtilis	*ComK*	van Sinderen *et al.*, 1994	15 (Nanamiya *et al.*, 2003)	4 (Hamoen *et al.*, 1998)	Maamar *et al.*, 2007
S. cerevisiae	*PDR3*	Delahodde *et al.*, 1995	51 (Belle *et al.*, 2006)	2 (Delahodde *et al.*, 1995)	Downstream readout PDR5 (Zenklusen *et al.*, 2008)
S. cerevisiae	*REB1*	Wang *et al.*, 1998	12 (Belle *et al.*, 2006)	3 (Wang *et al.*, 1998)	N/A
C. elegans	*ELT-2*	Fukushige *et al.*, 1999	N/A	Multiple (Fukushige *et al.*, 1999)	protein (Fukushige *et al.*, 1998)
D.	*ftz*	Schier *et al.*,	7-40	6	protein

melanogaster		1992	(Kellerman *et al.*, 1990)	(Schier *et al.*, 1992)	(Surkova *et al.*, 2008)
Mammals	*Nanog*	Boyer *et al.*, 2005	90 (*Oct4*) (Saxe *et al.*, 2009)	N/A	protein (Kalmar *et al.*, 2009)
Mammals	*Ets-1*	Majerus *et al.*, 1992	70-80 (Nishida *et al.*, 2007)	3 (Majerus *et al.*, 1992)	protein (De la Houssaye *et al.*, 2008)
Mammals	*c-Jun*	Angel *et al.*, 1988	150 (Bhoumik *et al.*, 2004)	2 (Angel *et al.*, 1988)	N/A

N/A – not available

a) Also indirect autoregulation through *Oct4-Sox2*

4.6 References

Acar M, Becskei A & van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435: 228-232.

Ajo-Franklin CM, *et al* (2007) Rational design of memory in eukaryotic cells. *Genes Dev* 21: 2271-2276.

Alon U (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits,* ((CRC) Press, Boca Raton, Fl), pp 320.

Altschuler SJ, Angenent SB, Wang Y & Wu LF (2008) On the spontaneous emergence of cell polarity. *Nature* 454: 886-889.

Angel P, Hattori K, Smeal T & Karin M (1988) The jun proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell* 55: 875-885.

Bar-Even A, *et al* (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* 38: 636-643.

Becskei A, Boselli MG & van Oudenaarden A (2004) Amplitude control of cell-cycle waves by nuclear import. *Nat Cell Biol* 6: 451-457.

Becskei A, Seraphin B & Serrano L (2001) Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO J* 20: 2528-2535.

Belle A, Tanay A, Bitincka L, Shamir R & O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* 103: 13004-13009.

Bhoumik A, Jones N & Ronai Z (2004) Transcriptional switch by activating transcription factor 2-derived peptide sensitizes melanoma cells to apoptosis and inhibits their tumorigenicity. *Proc Natl Acad Sci U S A* 101: 4222-4227.

Bintu L, *et al* (2005) Transcriptional regulation by the numbers: Models. *Curr Opin Genet Dev* 15: 116-124.

Boyer LA, *et al* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947-956.

Colman-Lerner A, *et al* (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437: 699-706.

De la Houssaye G, *et al* (2008) ETS-1 and ETS-2 are upregulated in a transgenic mouse model of pigmented ocular neoplasm. *Mol Vis* 14: 1912-1928.

Delahodde A, Delaveau T & Jacq C (1995) Positive autoregulation of the yeast transcription factor Pdr3p, which is involved in control of drug resistance. *Mol Cell Biol* 15: 4043-4051.

Elowitz MB, Levine AJ, Siggia ED & Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297: 1183-1186.

Friedman N, Cai L & Xie XS (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* 97: 168302.

Fukushige T, Hawkins MG & McGhee JD (1998) The GATA-factor elt-2 is essential for formation of the caenorhabditis elegans intestine. *Dev Biol* 198: 286-302.

Fukushige T, Hendzel MJ, Bazett-Jones DP & McGhee JD (1999) Direct visualization of the elt-2 gut-specific GATA factor binding to a target promoter inside the living caenorhabditis elegans embryo. *Proc Natl Acad Sci U S A* 96: 11883-11888.

Gari E, Piedrafita L, Aldea M & Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in saccharomyces cerevisiae. *Yeast* 13: 837-848.

Ghaemmaghami S, *et al* (2003) Global analysis of protein expression in yeast. *Nature* 425: 737-741.

Gossen M & Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* 89: 5547-5551.

Hamoen LW, Van Werkhoven AF, Bijlsma JJ, Dubnau D & Venema G (1998) The competence transcription factor of bacillus subtilis recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev* 12: 1539-1550.

Henssler EM, Bertram R, Wisshak S & Hillen W (2005) Tet repressor mutants with altered effector binding and allostery. *FEBS J* 272: 4487-4496.

Ingolia NT & Murray AW (2007) Positive-feedback loops as a flexible biological module. *Curr Biol* 17: 668-677.

Isaacs FJ, Hasty J, Cantor CR & Collins JJ (2003) Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci U S A* 100: 7714-7719.

Kalmar T, *et al* (2009) Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* 7: e1000149.

Karmakar R & Bose I (2007) Positive feedback, stochasticity and genetic competence. *Phys Biol* 4: 29-37.

Karmakar R & Bose I (2004) Graded and binary responses in stochastic gene expression. *Phys Biol* 1: 197-204.

Keller AD (1995) Model genetic circuits encoding autoregulatory transcription factors. *J Theor Biol* 172: 169-185.

Kosugi S, Hasebe M, Tomita M & Yanagawa H (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* 106: 10171-10176.

Maamar H, Raj A & Dubnau D (2007) Noise in gene expression determines cell fate in bacillus subtilis. *Science* 317: 526-529.

Maheshri N & O'Shea EK (2007) Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* 36: 413-434.

Majerus MA, Bibollet-Ruche F, Telliez JB, Wasylyk B & Bailleul B (1992) Serum, AP-1 and ets-1 stimulate the human ets-1 promoter. *Nucleic Acids Res* 20: 2699-2703.

Miller JA & Widom J (2003) Collaborative competition mechanism for gene activation in vivo. *Mol Cell Biol* 23: 1623-1632.

Murphy KF, Balazsi G & Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A* 104: 12726-12731.

Nanamiya H, *et al* (2003) Involvement of ClpX protein in the post-transcriptional regulation of a competence specific transcription factor, ComK protein, of bacillus subtilis. *J Biochem* 133: 295-302.

Newman JR, *et al* (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* 441: 840-846.

Nishida T, Terashima M, Fukami K & Yamada Y (2007) PIASy controls ubiquitination-dependent proteasomal degradation of ets-1. *Biochem J* 405: 481-488.

Ovechkina Y, *et al* (2003) Spindle formation in aspergillus is coupled to tubulin movement into the nucleus. *Mol Biol Cell* 14: 2192-2200.

Raj A, Peskin CS, Tranchina D, Vargas DY & Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4:

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A & Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 5: 877-879.

Raser JM & O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304: 1811-1814.

Salghetti SE, Caudy AA, Chenoweth JG & Tansey WP (2001) Regulation of transcriptional activation domain function by ubiquitin. *Science* 293: 1651-1653.

Samoilov M, Plyasunov S & Arkin AP (2005) Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proc Natl Acad Sci U S A* 102: 2310-2315.

Saxe JP, Tomilin A, Scholer HR, Plath K & Huang J (2009) Post-translational regulation of Oct4 transcriptional activity. *PLoS One* 4: e4467.

Schier AF & Gehring WJ (1992) Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. *Nature* 356: 804-807.

Surkova S, *et al* (2008) Characterization of the drosophila segment determination morphome. *Dev Biol* 313: 844-862.

Swain PS, Elowitz MB & Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* 99: 12795-12800.

Tan C, Marguet P & You L (2009) Emergent bistability by a growth-modulating positive feedback circuit. *Nat Chem Biol* 5: 842-848.

van Sinderen D & Venema G (1994) comK acts as an autoregulatory control switch in the signal transduction route to competence in bacillus subtilis. *J Bacteriol* 176: 5762-5770.

Wang KL & Warner JR (1998) Positive and negative autoregulation of REB1 transcription in saccharomyces cerevisiae. *Mol Cell Biol* 18: 4368-4376.

Weinberger LS, Burnett JC, Toettcher JE, Arkin AP & Schaffer DV (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 tat fluctuations drive phenotypic diversity. *Cell* 122: 169-182.

Weinberger LS, Dar RD & Simpson ML (2008) Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet* 40: 466-470.

Weinberger LS & Shenk T (2007) An HIV feedback resistor: Auto-regulatory circuit deactivator and noise buffer. *PLoS Biol* 5: e9.

Xiong W & Ferrell JE,Jr (2003) A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* 426: 460-465.

Zenklusen D, Larson DR & Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* 15: 1263-1271.

5 Future directions

5.1 Overview

We exploited spontaneous fluctuations in gene expression to describe the molecular mechanisms and kinetic pathways underlying transcriptional bursting. By quantifying the effect of activators on the dynamics of mRNA production, we were able to describe how activators selectively utilize different promoter elements to control the kinetics of initiation and reinitiation. However, this work only represents a small contribution to a much larger picture of transcriptional regulation. The transcriptional machinery has evolved to be complex and redundant to allow integrations of multiple signaling pathways and provide biological robustness. In eukaryotes, an additional complexity lies in the fact that protein binding sites are often wrapped in nucleosomes. How do transcription factors, nucleosomes, and the underlying DNA sequence act together to determine the kinetics of RNA Pol II transcription? Although there is still debate about the relative contribution of DNA sequence preference and chromatin remodeling factors in determining nucleosome positions (Jiang and Pugh, 2009), the universal nature of the nucleosome-positioning code has been established. It is widely accepted that both individual nucleosome locations and higher-order chromatin structure are explicitly encoded in the genome (Kaplan *et al.*, 2009). The most common view for how nucleosome positioning regulates gene expression is via controlling the accessibility of the core promoter. It would be interesting to see how promoter accessibility influences the transcriptional dynamics. In Section 5.2, we describe our attempts to characterize the

effects of nucleosome positioning on gene expression noise and uncover the mechanism by which such effects are mediated.

Defining the detailed kinetic pathway by which activators mediate the steps involved in PIC formation remain a major scientific challenge. Although our approach provides a coarse-grained view of how activators regulate the rates of initiation and reinitiation, it does not reveal how activators modulate the formation and function of the biochemical intermediates involved in these steps. These biochemical intermediates often involve additional transcriptional coactivators, such as the Mediator complex. In order to propose a more detailed kinetic pathway of transcriptional activation, we need to better define the kinetic roles of coactivators. In Section 5.3, we describe our initial efforts at determining the kinetic roles of the Mediator by direct recruitment to the promoter.

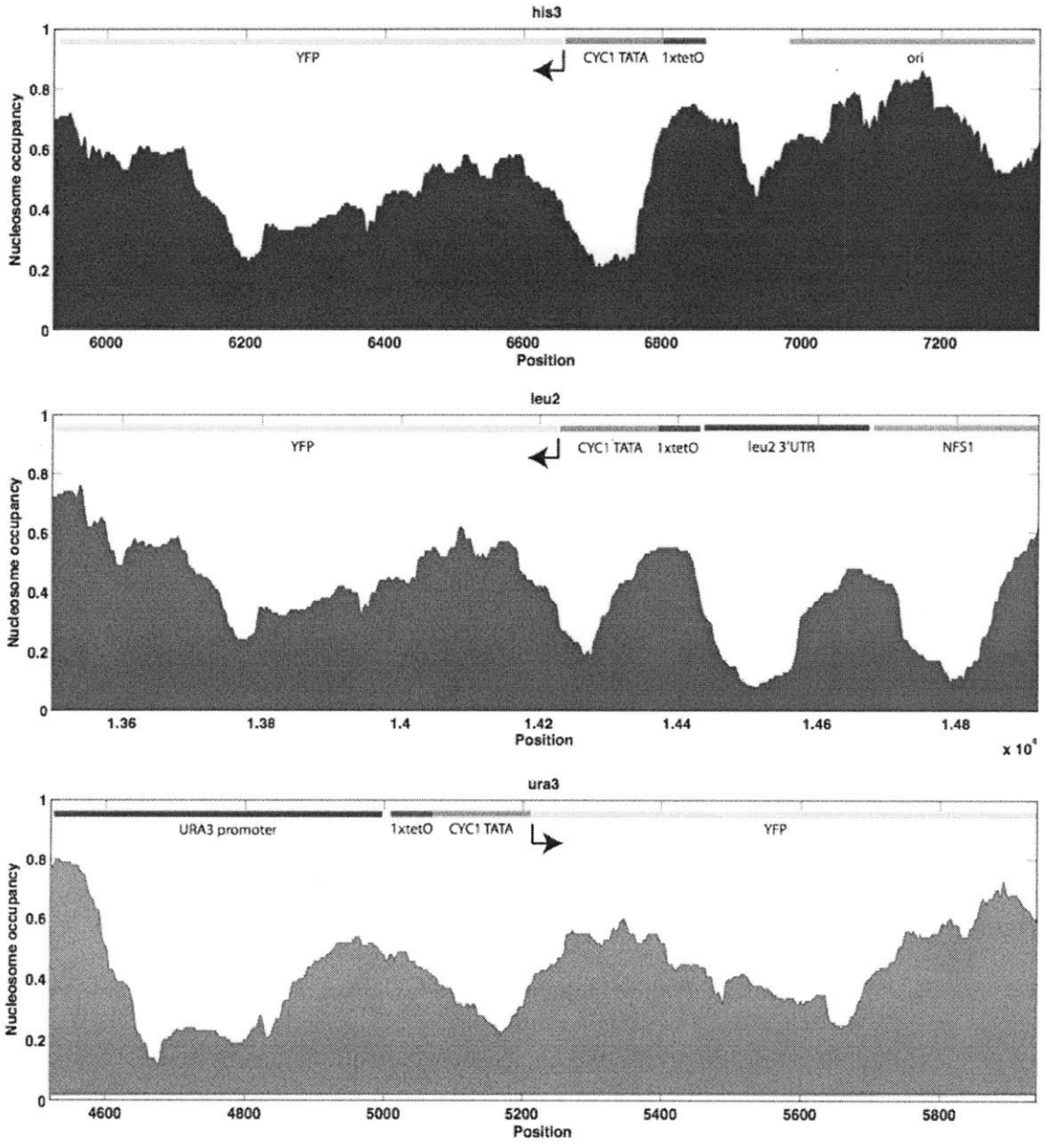
At the network level, transcriptional regulation alone offers a rich set of dynamical behaviors, spanning multistability, oscillation and noise suppression. These irregular dynamic behaviors have important implications for human disease and developmental programs (Maheshri and O'Shea, 2007; Raj and van Oudenaarden, 2008). As demonstrated in Chapter 4, relatively small changes in the promoter region can have large effects on the network dynamics. Therefore, the molecular mechanisms of transcription regulation constitute an important knowledge since disturbances in transcriptional programs can have profound physiological consequences. In Section 5.4, we describe our efforts in further connecting molecular level changes in the promoter to the emergence of bimodal gene expression, a hallmark in cellular decision-making.

5.2 Defining the effect of nucleosome positioning on the kinetics of transcription

While the strong effect of chromosome positioning on gene expression noise has been previously identified (Becskei *et al.*, 2005; Batuda *et al.*, 2007), little is known about the effects of local nucleosome positioning on transcriptional noise. Strikingly, we observed substantial differences in mean expression and variability when the same 1xtetO promoter is placed at different loci using different integrating constructs. The promoter variants at *leu2* and *ura3* have higher dynamic ranges of expression than their *his3* counterpart (Figure 3.5). The burst statistics reveal that for the *leu2* and *ura3* promoter variants, the basal burst size is lower while the maximum burst size is higher. The burst frequency does not vary significantly among these promoter variants. We concluded from this preliminary finding that nucleosome positioning affects how burst size is regulated by activators.

To determine if burst size regulation was correlated to nucleosome positioning or accessibility, we applied an *in silico* nucleosomal occupancy model (Kaplan *et al.*, 2009) on our promoter variants. For the *leu2* and *ura3* variants, the *in silico* model suggests the region upstream of the tetO binding site is nucleosome-free but the tetO binding site, and part of the *CYC1* core promoter lie in a nucleosome-occluded region (Figure 5.1). For the *his3* variant, on the other hand, the *CYC1* core promoter is relatively nucleosome-free. The nucleosome-free region immediately upstream of tetO in the *leu2* and *ura3* variants is likely caused by the flanking sequences of the integration construct, rather than an intrinsic property of those loci. Indeed, the *in silico* model predicts a similar upstream nucleosome free region at the *his3* locus when an *ADH1* terminator sequence is upstream

of the 1xtetO site or the 1xtetO is replaced with 7xtetO (7 tetO binding sites in tandem). We experimentally measured the burst statistics of these two promoter variants (see Chapter 4 for the results of 7xtetO), and obtained results similar to the *leu2* and *ura3* variants. Taken together, a nucleosome-free region upstream of the binding site leads to a larger dynamic range of burst size regulation.



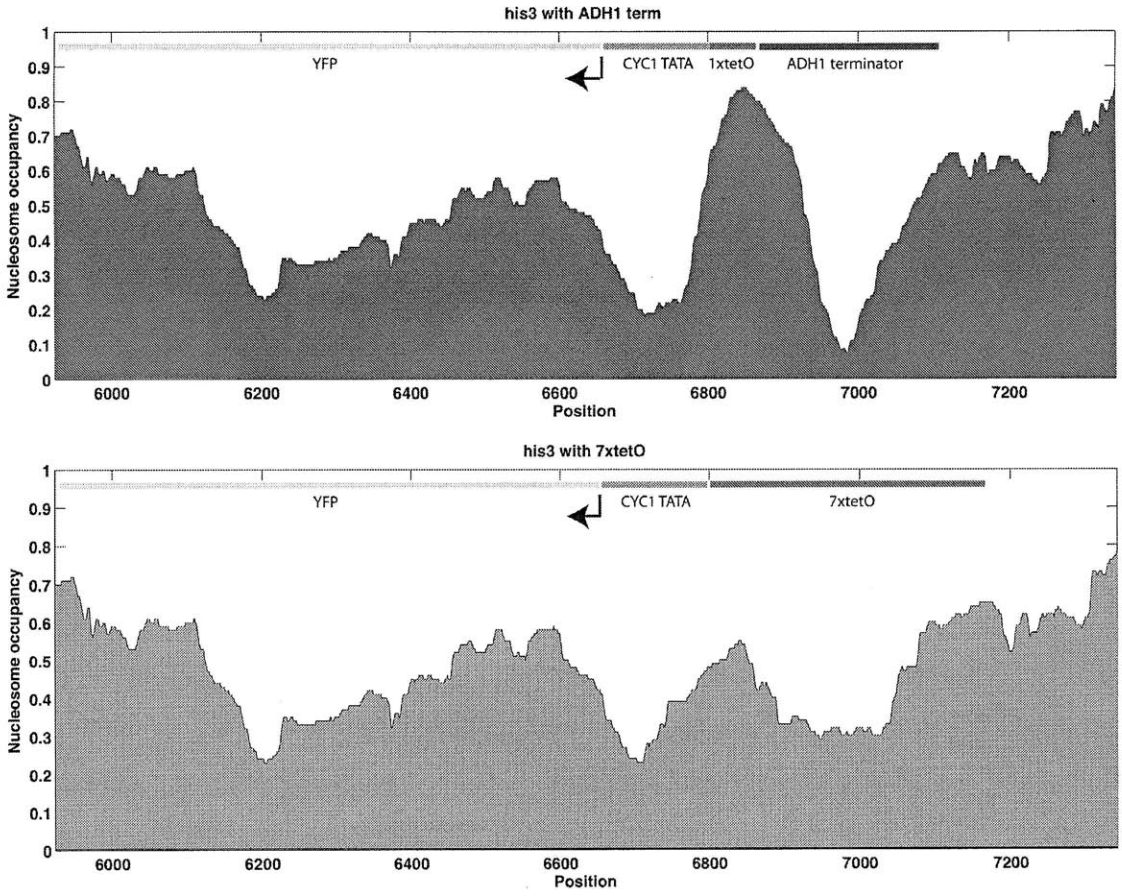


Figure 5.1. Predicted nucleosome occupancy of the tetO promoter variants. The maps of nucleosome occupancy were obtained using the computational model by Kaplan *et al.* (2009). The nucleosome occupancy on the *CYC1* core promoter increases when nucleosome-excluding sequence is placed upstream of the tetO site.

Interestingly, by introducing a nucleosome–excluding sequence upstream of the promoter region, we could increase the dynamic range and variability of gene expression significantly. Our findings lend support to the idea that statistical positioning of nucleosomes can directly affect the kinetic roles of activators and hence the heterogeneity of gene expression. Here, by preventing nucleosome formation, the upstream sequence may confer higher equilibrium nucleosome occupancy on the core promoter by

repositioning the nucleosomes in the region thereby leading to lower basal expression. However, the mechanism by which this sequence upregulates reinitiation at high activator levels is unclear. One explanation may rely on the regional accessibility of the promoter. By preventing nucleosome formation, the upstream sequence may create a more open chromatin state within the promoter that is permissive to reinitiation events. An obvious future direction then is to test these possibilities. Initially, we will test how far an anti-nucleosomal sequence can be from the activator binding site and still has an effect. To test the regional accessibility hypothesis, the chromatin state of the region can be assayed by ChIP.

5.3 Probing the kinetic roles of the Mediator by direct recruitment

We lack complete knowledge of the biochemical intermediates that occur during the slow initiation step that governs burst frequency and how activators regulate these intermediates. Since both TBP and RNA Pol II are present at the *CYC1* core promoter in repressed conditions (Martens *et al.*, 2001), RNA Pol II recruitment is not rate-limiting for our promoters. One possibility is the rate-determining step may depend on SAGA and its composition; gene-specific activators may control the utility of the SAGA pathway. However, a recent study (Lee *et al.*, 2010) shows that SAGA occupies the *CYC1* promoter under both repressed and induced conditions, whereas Mediator is present only under induced conditions. Because SAGA is not required for activation at the *CYC1* promoter, we hypothesized that the main regulatory role of the VP16 activation domain in tTA is to recruit a certain form of the Mediator. Indeed, recruitment of Mediator can be rate-limiting in many contexts (Weake and Workman, 2010).

To demonstrate that recruitment of Mediator (but not SAGA) is sufficient to induce gene expression, we created a tetR-Med2p fusion protein. Med2p is found in the tail domain of the Mediator complex and it is known to be a potent activator by itself (Chang and Ptashne, 2004; Wang *et al.*, 2010). By directly recruiting the tetR-Med2p fusion protein to the tetO-*CYC1* core promoter, we can circumvent the need of an activator for gene activation. Moreover, the Med2p fusion protein may utilize canonical and non-canonical TATA elements differently because of possible changes in the conformation and stability of the resulting transcription complexes. Indeed, titrations of tetR-Med2p lead to markedly different dose-response curves (Figure 5.2). Surprisingly, the wild-type 1xtetO promoter remains fully activated oven a large range of doxycycline concentrations corresponding to a ~100-fold range of tetR-Med2 levels. Moreover, the dynamic range of expression increases for a number of tetO and TATA mutants, implying that burst size may be up-regulated. One possibility that explains these findings is that since tetR-Med2p makes direct contact with the DNA, it does not unbind as easily as when brought to the promoter via interaction with an activator. The prolonged effective residence time of Mediator then leads to a longer life-time of the transcription complex. As a result, the saturation of the canonical element occurs at a low level of activator (in this case tetR-Med2) since the life-time of the reinitiation complex is intrinsically longer. This model predicts that the burst size attains a maximum at a low level of tetR-Med2p, and that tetR-Med2p enables TATA and binding site mutants to realize a higher maximum burst sizes. Moreover, since Mediator recruitment could be rate-limiting, direct recruitment of Mediator to the canonical TATA may also increase to burst frequency. These predictions can be tested by mRNA FISH measurements.

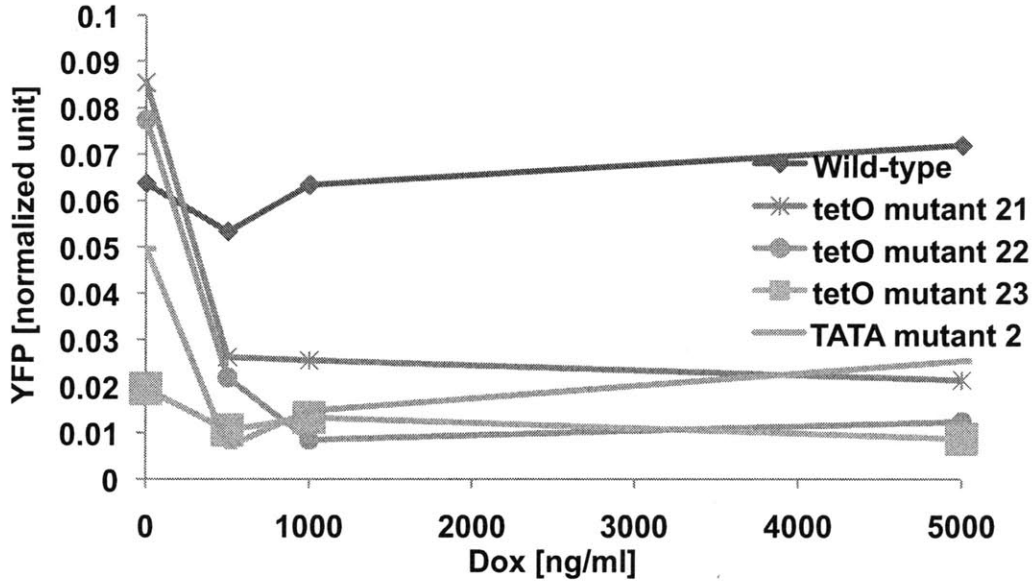


Figure 5.2. Mean YFP expression of 1xtetO promoter variants under the control of tetR-Med2p. An *ADH1* promoter was used to constitutively express tetR-Med2p. The active tetR-Med2p level was modulated by doxycycline. All promoter variants are integrated at the *his3* locus.

5.4 Bridging molecular details and functional consequences: tuning of bimodality in transcriptional positive feedback loops

Most previous investigations of gene expression noise, including our own described in Chapter 4, have found or assumed the characteristic signature of burst frequency regulation by transcription regulators. However, our findings in Chapter 3 strongly suggest that transcription activators can regulate both burst size and burst frequency, and they have broad implications for functions. As demonstrated in Chapter 4, transcriptional positive feedback loops can lead to bimodal gene expression and provide a mechanistic ground to cellular decision-making. How activators modulate

burst size and frequency regulation can have important outcomes when the positive feedback loop is involved.

It has been previously shown the deterministic bistable states in a positive feedback loop can be stabilized by burst frequency regulation (Friedman *et al.*, 2006; Carmaker and Bose, 2007). In contrast, theoretical and computational investigations (Quinn and Maheshri, unpublished results) demonstrate that burst size regulation has the opposite effect of destabilizing deterministic bistable states, and that mixed burst frequency and size regulation can lead to very different outcomes depending on the relative amount of regulation between burst size and burst frequency. It would be interesting to see if these model predictions can be realized in our synthetic positive feedback system. For example, we can use the Type II TATA mutant (Section 3.3.1) to create a positive feedback circuit with primarily burst size regulation. By relating molecular level changes to profound functional outcomes on decision making, we can gain a deeper understanding on how tunability and plasticity of gene regulatory networks underlie the evolvability and survival strategies of an organism, and thereby achieve a highly mechanistic genotype-phenotype map. Such pursuit is, beyond any doubt, a grand goal of systems biology.

5.5 References

Batada NN & Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* 39: 945-949.

Becskei A, Kaufmann BB & van Oudenaarden A (2005) Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat Genet* 37: 937-944.

Cheng JX, Gandolfi M & Ptashne M (2004) Activation of the Gal1 gene of yeast by pairs of 'non-classical' activators. *Curr Biol* 14: 1675-1679.

Friedman N, Cai L & Xie XS (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* 97: 168302.

Jiang C & Pugh BF (2009) Nucleosome positioning and gene regulation: Advances through genomics. *Nat Rev Genet* 10: 161-172.

Kaplan N, *et al* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458: 362-366.

Karmakar R & Bose I (2007) Positive feedback, stochasticity and genetic competence. *Phys Biol* 4: 29-37.

Lee SK, *et al* (2010) Activation of a poised RNAPII-dependent promoter requires both SAGA and mediator. *Genetics* 184: 659-672.

Maheshri N & O'Shea EK (2007) Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* 36: 413-434.

Martens C, Krett B & Laybourn PJ (2001) RNA polymerase II and TBP occupy the repressed CYC1 promoter. *Mol Microbiol* 40: 1009-1019.

Pugh BF (2000) Control of gene expression through regulation of the TATA-binding protein. *Gene* 255: 1-14.

Raj A & van Oudenaarden A (2008) Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135: 216-226.

Roeder RG (2005) Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* 579: 909-915.

Segal E & Widom J (2009) What controls nucleosome positions?. *Trends Genet* 25: 335-343.

Wang X, Muratani M, Tansey WP & Ptashne M (2010) Proteolytic instability and the action of nonclassical transcriptional activators. *Curr Biol* 20: 868-871.

Weake VM & Workman JL (2010) Inducible gene expression: Diverse regulatory mechanisms. *Nat Rev Genet* 11: 426-437.

Appendices

Appendix 1 Yeast strains and plasmids

Appendix 1.1 Yeast strains

Strain Name	Relevant Genotype
Y88	MATa *ura3Δ::*P_{1xtetO}-*YFP-HIS3Mx6*
Y123	MATa *leu2Δ::*$P_{Adh1term-7xtetO}$-*YFP-HIS3*
Y124	MATa *leu2Δ::*P_{1xtetO}-*YFP-HIS3*
Y128	MATa *his3::*P_{1xtetO}-*tTA-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y139	MATa *his3::*P_{1xtetO}-*YFP-HIS3*
Y139M	MATa/α *his3/his3::*P_{1xtetO}-*YFP-HIS3/* P_{1xtetO}-*CFP-HIS3*
Y145	MATa *his3::*P_{1xtetO}-*tTA-CFP-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y161	MATa *his3::*P_{7xtetO}-*tTA-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y162	MATa *his3::*P_{7xtetO}-*tTA-CFP-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y163	MATa *his3::*P_{7xtetO}-*YFP-HIS3*
Y163M	MATa/α *his3/his3::*P_{7xtetO}-*YFP-HIS3/* P_{7xtetO}-*CFP-HIS3*
Y253	MATa *met4Δ::TRP1 met30Δ::LEU2 his3::* P_{1xtetO}-*Ub-tTA-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y254	MATa *met4Δ::TRP1 met30Δ::LEU2 his3::* P_{7xtetO}-*Ub-tTA-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y286	MATa *his3::*$P_{Adh1term-1xtetO}$-*YFP-HIS3*
Y290	MATa *his3::*P_{1xtetO}-*NLS-tTA-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y291	MATa *his3::*P_{7xtetO}-*NLS-tTA-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y293	MATa *his3::*$P_{Adh1term-1xtetO}$-*tTA-HIS3 leu2Δ::*$P_{Adh1term7xtetO}$-*YFP*
Y349	MATa *his3::*$P_{1xtetO(Mutant\ 12)}$-*YFP-HIS3*
Y350	MATa *his3::*$P_{1xtetO(Mutant\ 17)}$-*YFP-HIS3*
Y352	MATa *his3::*$P_{1xtetO(Mutant\ 21)}$-*YFP-HIS3*
Y353	MATa *his3::*$P_{1xtetO(Mutant\ 22)}$ -*YFP-HIS3*

Y354	MATa *his3::*$P_{1xtetO(Mutant\ 23)}$-*YFP-HIS3*
Y355	MATa *his3::*$P_{1xtetO(TATA\ mutant\ 1)}$-*YFP-HIS3*
Y356	MATa *his3::*$P_{1xtetO(TATA\ mutant\ 2)}$-*YFP-HIS3*
Y487	MATa *his3::*$P_{1xtetO(TATA\ no\ weak\ sites)}$-*YFP-HIS3*
Y488	MATa *his3::*$P_{1xtetO(TATA\ mutant\ 3)}$-*YFP-HIS3*
Y500	MATa *leu2Δ::*$P_{1xtetO(TATA\ mutant\ 1)}$-*YFP-HIS3*
Y501	MATa *leu2Δ::*$P_{1xtetO(TATA\ mutant\ 2)}$-*YFP-HIS3*
Y502	MATa *leu2Δ::*$P_{1xtetO(TATA\ no\ weak\ site)}$-*YFP-HIS3*

All *S. cerevisiae* strains were derived from the W303 background. All transformations were carried out using the LiAC/ssDNA/PEG method (Guthrie *et al.*, 2004)

Appendix 1.2 Yeast plasmids

Plasmid Name	Description
B56	Yeast centromeric P_{ADH1}-*tTA* P_{1xtetO}-*YFP* P_{PGK1}-*tdTomato*
B059Y	Yeast centromeric P_{1xtetO}-*YFP*
B117Y	Yeast centromeric P_{7xtetO}-*YFP*
B126	Yeast centromeric P_{GAL1}-*tTA*
B172	Yeast centromeric P_{GAL1}-*tTA-CFP*
B179	Yeast centromeric P_{ADH1}-*tTA-CFP*
B179b	Yeast centromeric P_{ADH1}-*tTA-YFP*
B228	Yeast centromeric P_{ADH1}-*tTA*
B314	Yeast centromeric P_{ADH1}-*NLS-tTA*
B317	Yeast centromeric P_{ADH1}-*NLS-tTA-YFP*
B328	Yeast centromeric P_{ADH1}-*Ub-tTA*
B329	Yeast centromeric P_{GAL1}-*Ub-tTA*
B499	Yeast centromeric P_{ADH1}-*NLS-tetR-Med2*

All yeast centromeric plasmids used in this study were based on pCM189 (Gari *et al.*, 1997)

Appendix 1.3 References

Gari E., L. Piedrafita, M. Aldea, E. Herrero, *Yeast* **13**, 837 (1997).

Guthrie C., G. R. Fink, *Guide to yeast genetics and molecular and cell biology. Part A* (Elsevier Academic Press, Amsterdam, 2004).

Appendix 2 Protocols

Appendix 2.1 Growth conditions

Yeast cultures were grown at 30°C in 96-well deep-well plates on a Variomag® Teleshake electronic stirrer (Thermo Scientific) at a speed of 1200 rpm/minute in synthetic minimal medium supplemented with the appropriate sugar and amino acids. For doxycycline titration experiments, a freshly grown overnight culture was diluted to an A_{600nm} of 0.05 and induced with doxycycline (Sigma-Aldrich), then grown for 12 hours to reach A_{600nm} of 1. Cells were further diluted to an A_{600nm} of 0.1 and grown an additional 8-12 hours to ensure steady-state fluorescence expression. For galactose induction experiments, freshly grown cells were diluted to an A_{600nm} of 0.05 and grown in synthetic dextrose medium to an A_{600nm} of 1, then transferred to synthetic medium containing galactose from 0 to 2%, balanced by raffinose, such that the total sugar content remained at 2%. Cells in galactose/raffinose media were grown an additional 8 hours before the measurement of fluorescence expression.

Appendix 2.2 Fluorescence measurements

Population-averaged fluorescence intensities were evaluated with a Varioskan® Flash spectrofluorimeter (Thermo Fisher Scientific). Single cell fluorescence measurements were taken using either flow cytometry, using an HTS equipped LSRII flow cytometer (BD Biosciences), or fluorescence microscopy. Microscopic studies were performed with a Zeiss AxioObserver inverted microscope and Metamorph data acquisition software (Universal Imaging). For each sample, 700-2000 cells were imaged. Average cell fluorescence was obtained by dividing total cell fluorescence by cell area. For all fluorescence measurements the cellular autofluorescence was measured using a W303 strain without fluorescent reporters. For all experiments performed, W303 strains carrying integrated copies of *ADH1*Pr-XFP at the *leu2* locus were always used as positive controls to account for any changes in illumination intensity and/or growth conditions. These fluctuations were small, less than 10% of the expression level. After correcting for autofluorescence, all population-averaged fluorescence measurements were normalized by using the corresponding positive control value.

Appendix 2.3 Fluorescence in situ hybridization (FISH) for detecting single mRNA molecules

Oligonucleotide probes targeting YFP and tetR transcripts (Table A2.1) were designed using the web-based probe designer at http://www.singlemoleculefish.com and obtained from Biosearch Technologies, Inc. An additional amine group was introduced at the 3' end of the oligonucleotide for coupling to the tetramethylrhodamine (TMR) fluorophore. Probes were purified on an HPLC column and only the highly TMR-coupled

probes were retained. Fixation, hybridization and washing were performed as described in (Raj *et al.* 2008). The hybridization solution and wash buffer contained 10% formamide. Oligonucleotide probes at ~5 μM were diluted 50-fold into the hybridization solution. Imaging was performed on a Zeiss AxioObserver inverted microscope equipped with an X-Cite® 120 PC mercury arc lamp (EXFO), a 100X/1.40 objective (Zeiss) and a rhodamine-specific filter set (Chroma Technology Cat. No. 31000v2).

Table A2.1. Oligonucleotide probes used in fluorescence *in situ* hybridization

Sequence	Probe Name
AATTCTTCACCTTTAGACAT	YFP01
AATTGGGACAACACCAGTGA	YFP02
CATCACCATCTAATTCAACC	YFP03
GACAGAAAATTTGTGACCAT	YFP04
CATCACCTTCACCTTCACCG	YFP05
AAGGTCAATTTACCGTAAGT	YFP06
ACCAGTAGTACAAATTAATT	YFP07
GTTGGCCATGGAACTGGCAA	YFP08
ATAACCTAAAGTAGTGACTA	YFP09
ATCTAGCAAAACACATTAAA	YFP10
TGTTGTTTCATATGATCTGG	YFP11
CATGGCAGACTTGAAAAAGT	YFP12
CTTTCTTGAACATAACCTTC	YFP13
GTCATCTTTGAAAAAAATAG	YFP14
CAGCTCTGGTCTTGTAGTTA	YFP15
GTATCACCTTCAAACTTGAC	YFP16
TAATTCGATTCTATTAACTA	YFP17
TCTTCTTTAAAATCAATACC	YFP18
TTTGTGACCTAAAATGTTAC	YFP19

GAGAGTTATAGTTGTATTCC	YFP20
TCAGCAGTGATGTAAACATT	YFP21
CTTTGATACCATTCTTTTGT	YFP22
TTGTGTCTAATTTTGAAGTT	YFP23
TTGAACACCACCATCTTCAA	YFP24
TTTGTTGATAATGGTCAGCT	YFP25
GGACCATCACCAATTGGAGT	YFP26
AATGGTTGTCTGGTAACAAG	YFP27
AAGGCAGATTGATAGGATAA	YFP28
CTTTTCGTTTGGATCTTTGG	YFP29
CTAACAAGACCATGTGGTCT	YFP30
ATACCAGCAGCAGTAACAAA	YFP31
ACAATTCATCCATACCATGG	YFP32
CTTTTATCTAATCTAGACAT	tetR_01
CTAATGCGCTGTTAATCACT	tetR_02
ATTCCGACCTCATTAAGCAG	tetR_03
GTTTACGGGTTGTTAAACCT	tetR_04
TCTACACCTAGCTTCTGGGC	tetR_05
ATGCCAATACAATGTAGGCT	tetR_06
GCAAAGCCCGCTTATTTTTT	tetR_07
ATCTCAATGGCTAAGGCGTC	tetR_08
GTGAGTATGGTGCCTATCTA	tetR_09
TTTCCCCTTCTAAAGGGCAA	tetR_10
TTACGTAAAAAATCTTGCCA	tetR_11
CACATCTAAAACTTTTAGCG	tetR_12
CCATCGCGATGACTTAGTAA	tetR_13
TGTACCTAAATGTACTTTTG	tetR_14
CATACTGTTTTTCTGTAGGC	tetR_15
GGCTAATTGATTTTCGAGAG	tetR_16
GTGAAAAACCTTGTTGGCAT	tetR_17
AGTGCATATAATGCATTCTC	tetR_18

AAGTAAAATGCCCCACAGCG	tetR_19
TGATCTTCCAATACGCAACC	tetR_20
CTTCTTTAGCGACTTGATGC	tetR_21
TCAGTAGTAGGTGTTTCCCT	tetR_22
TCGTAATAATGGCGGCATAC	tetR_23
GATCAAATAATTCGATAGCT	tetR_24
AAGGCTGGCTCTGCACCTTG	tetR_25
GATCAATTCAAGGCCGAATA	tetR_26
GTTGTTTTTCTAATCCGCAT	tetR_27
GCGGACCCACTTTCACATTT	tetR_28

Appendix 2.4 Image acquisition and analysis for FISH

The sample was placed on a coverglass coated with poly-lysine to enhance adhesion. Z-stacks of 17 images 0.3 microns apart spanning the full cellular volume were taken for each sample. Well-defined signals from diffraction-limited spots were observed in cells with low levels of mRNA expression, with each spot representing a single mRNA molecule (Raj *et al*, 2008). The mRNA molecules in each cell were counted using custom software written in MATLAB (Mathworks). Briefly, the algorithm applied region-based thresholding, identified local maxima as potential spots, and counted the number of connected components in three dimensions. We could only reliably detect ~20 mRNA molecules per cell both manually and via the algorithm, because at higher levels signals were intense with large overlap between diffraction-limited spots. As such, we created calibration curves for the integrated rhodamine intensity versus the number of molecules obtained by the counting algorithm for test cells that had low expression levels, and obtained a linear correlation for up to 15 mRNA molecules. The simple linear

relationship allowed us to extrapolate the number of molecules from integrated fluorescence intensity in cells with large number of mRNA (i.e. > 15). Extrapolation was only used to obtain mRNA numbers if fluorescence signals were greater than the y-intercept of the calibration curve. The mRNA numbers from particle counting were used when the fluorescence signals fell below the y-intercept.

Appendix 2.5 Quantification of tTA half-life

Degradation rate of tTA was measured by *GAL1* promoter shutoff experiments followed by quantitative western blotting. Briefly, strains with tTA expression under a galactose–inducible promoter were grown in 0.5% galactose and 2% raffinose for 6 hours to allow tTA expression. Then, cells were transferred to 2% glucose to stop expression. Samples were collected and frozen at specific time points after the promoter shutoff. Cell lysates were prepared according to the protocol in (Belle *et al.* 2006). Western blotting was performed according to standard procedures. Blots were stained with a 1:2000 dilution of an affinity purified mouse monoclonal antibody raised against the wild type bacterial Tet repressor (MoBiTec Cat. No. TETO2). Membrane images were acquired by the AlphaImager® (Alpha Innotech).

The half-life of tTA variants was determined from Western blots as in (Belle *et al.* 2006). To generate densitometric data, images of the immunoblots were processed with ImageJ (NIH). The 8-bit image was converted to grayscale and inverted such that the dark areas (protein bands) became light. Background intensity was then subtracted from the image. A rectangular box was defined manually to enclose a single band completely, and the integrated density within the box was recorded. We observed a lag of ~60 minutes before the protein started to degrade for all tTA variants. This lag is likely due to

the time required for complete mRNA depletion and protein maturation. After the lag, protein levels based on densitometric data followed reasonably first-order decay kinetics.

Appendix 2.6 Quantification of mRNA half-life

To measure the degradation rate of mRNA, we grew a closed-loop strain expressing both tTA and a YFP reporter for 18 hours with dilution to ensure steady-state expression and then stopped transcription by adding 10 μg/mL doxycycline and 3 μg/mL transcriptional inhibitor thiolutin (Sigma-Aldrich). Cells were fixed at specific time points after inhibition. The abundance of both tTA and YFP mRNAs were measured by FISH as described above.

Appendix 2.7 Stochastic simulations

Direct stochastic simulations were performed using Dizzy http://magnet.systemsbiology.net/software/Dizzy/) (Ramsey *et al.s*, 2005)

Appendix 2.8 References

Belle A., A. Tanay, L. Bitincka, R. Shamir, E. K. O'Shea, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13004 (2006).

Raj A., P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, *Nat. Methods* **5**, 877 (2008).

Ramsey S., D. Orrell, H. Bolouri, *J. Bioinform Comput. Biol.* **3**, 415 (2005).

Appendix 3 A high-throughput, distribution-based approach to parameter estimation in deterministic models

Appendix 3.1 Introduction

In this section, we propose and test a methodology to rapidly and quantitatively characterize a library of genetic components. The concept is as follows: we define one component as a single gene (with promoter) and any gene-specific transcription factors. Its steady-state behavior is captured using a model defined by a small set of thermodynamic and kinetic parameters, such as the model in Section 4.2.1. Next, we create a library of this component by mutagenesis (e.g., of the promoter) - each member then has a potentially different set of parameters. Rather than looking at individual members, we quantitatively characterize the entire library with high-throughput single-cell measurements.

Appendix 3.2 Results and Discussion

To demonstrate this approach, we employed the Tet-OFF system (Gari *et al.*, 1997) and monitored single-cell gene expression state with YFP and flow cytometry. Because there is a distribution of responses even in a genetically homogeneous population, we used a constitutively expressed RFP reporter to account for non-genetic but correlated variation due to plasmid fluctuations and global noise. Both reporters were put on the same centromeric plasmid.

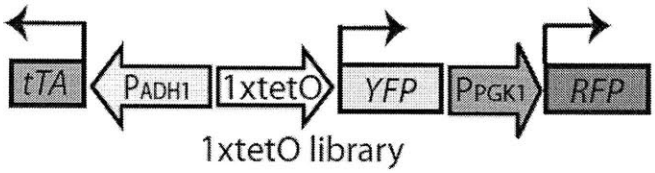


Figure A3.1. The one-component system. The system is composed of a constitutively expressed tTA activator, an 1xtetO promoter driving the expression of YFP, and a constitutively expressed RFP that serves as a global reporter. All three parts were placed on the same centromeric plasmid.

Microscopic analysis (Figure A3.2) of YFP expression from the one component indicates a bimodal expression profile. The high peak is likely due to the doubling of centromeric plasmid. The non-genetic variation effectively accounts for over 80% of the cell-to-cell variation since RFP is highly correlated with YFP in this case.

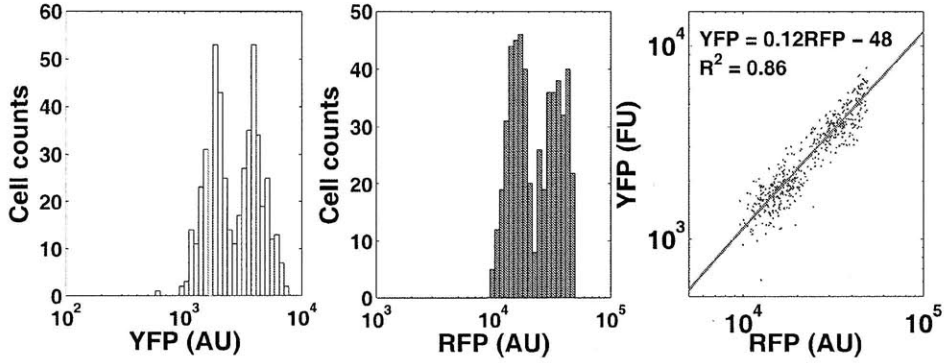


Figure A3.2. Correlated variation due to global noise and plasmid fluctuations. Effectively over 80% of the cell-to-cell variation can be accounted for by the RFP global reporter.

We designed two tetO libraries (diversity ~10^3), termed Library 3 and Library 4 (Figure A3.3), based on the PSAM in Section 3.2.5 (Figure 3.4). The distribution of dissociation constant K_D of the library can be predicted using the PSAM and equation

[3.7]. Figure A3.4 shows the K_D distribution of Libraries 3 and 4 predicted by the additive model. The K_D distribution spans two orders of magnitude. Figure A3.5 shows the density plots of steady-state YFP expression profiles at various doxycycline concentrations before (left) and after (right) correction with the global reporter. The noise in the wild type strain can be reduced by over 80% with the global reporter. In contrast, there is no significant noise reduction in Libraries 3 and 4 since genetics differences account for much of the variation.

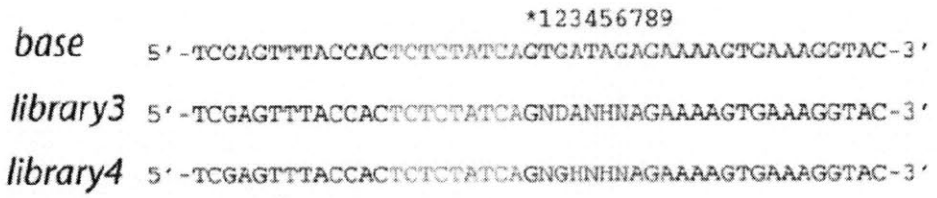


Figure A3.3. Design of tetO promoter mutant libraries. The design was guided by the position specific affinity matrix of tetO (Figure 3.4). Codes for degenerate bases: N – G,A,C,T; D – G,A,T; H – A,T,C.

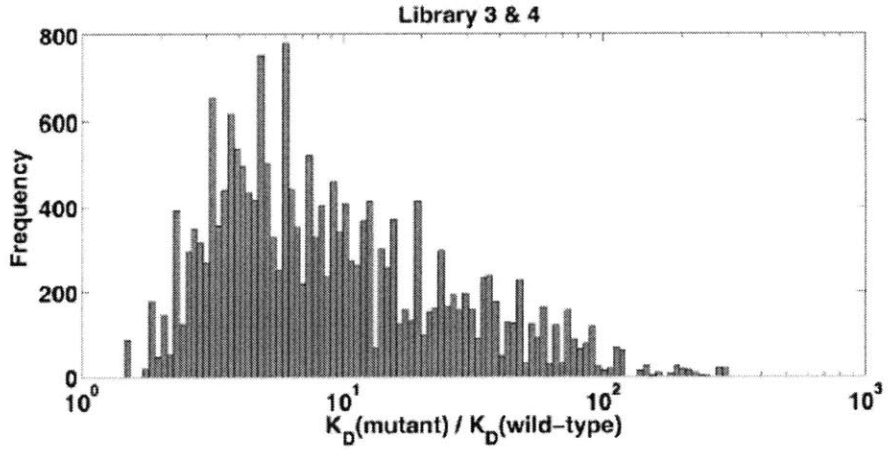


Figure A3.4 The K_D distributions of Libraries 3 and 4. The K_D value of each library member in Figure A3.3 was predicted using the PSAM (Figure 3.4) and the additive model (equation [3.7]).

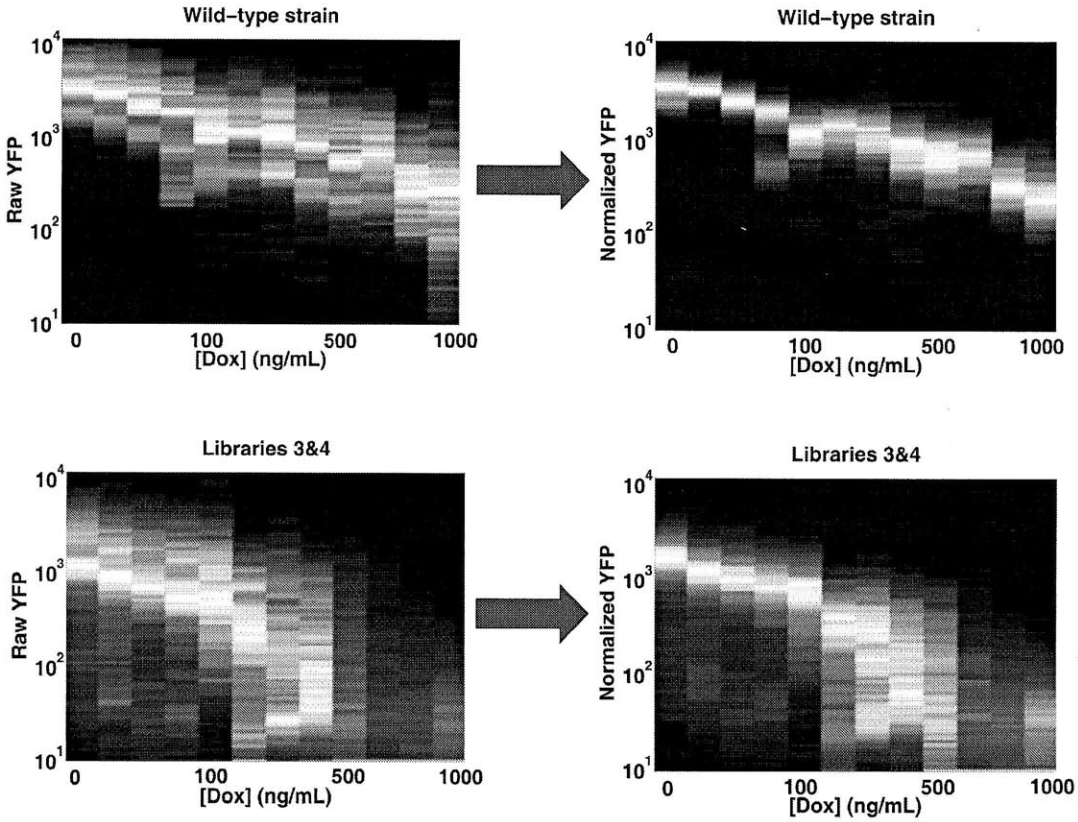


Figure A3.5. Density plots of steady-state YFP expression profiles of the mutant libraries. Shown are the profiles at various doxycycline concentrations before (left) and after (right) correction with the global reporter.

We looked at individual mutant clones to determine how mutations in the *tetO* sequence affected the regulatory response of the promoter. Figure A3.6 shows the dose response curves of the wild-type strain and four selected individual mutant clones along with fits (solid line) to the model in Section 4.2.1. The only one parameter varied across these clones is the dissociation constant K_D (K in Section 4.2.1). The good agreement between data and model confirms that the model can describe the steady-state expression of the mutant library with a distribution of K_D values. To rapidly obtain the K_D distribution, we performed dox titrations on the entire library, and back calculated the K_D

distribution using the model. We could not, however, differentiate K_D values greater than 25 since they corresponded to the off population in the absence of dox. Indeed, the library K_D distribution spans a much larger range (Figure A3.7) and agrees qualitatively to the K_D distribution predicted by the PSAM model (Figure A3.4).

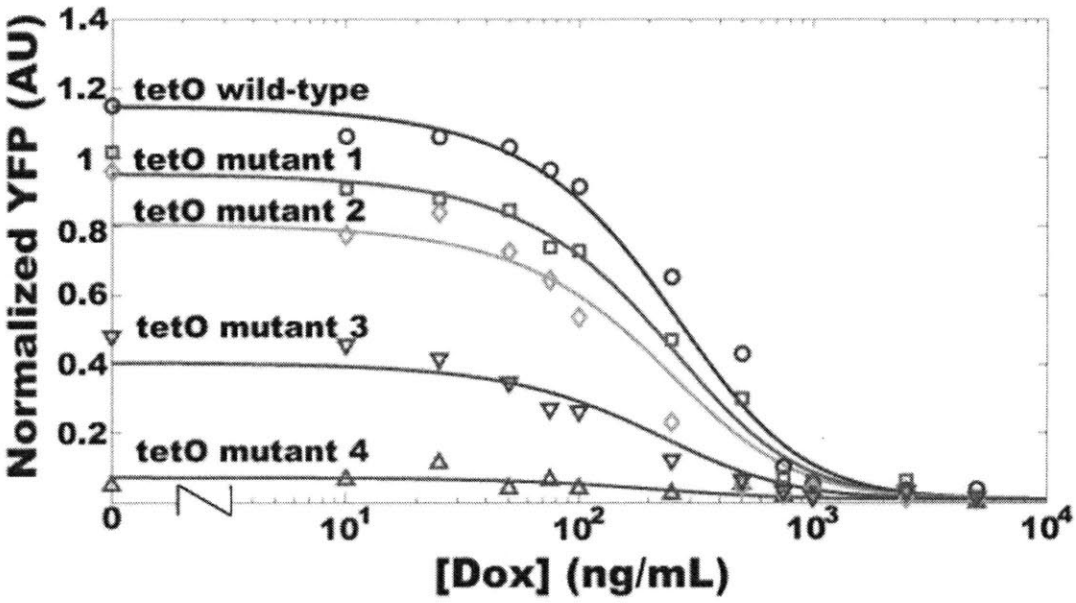


Figure A3.6. The dose response curves of four selected individual tetO mutants. The solid lines are fits to the model in Section 4.2.1.

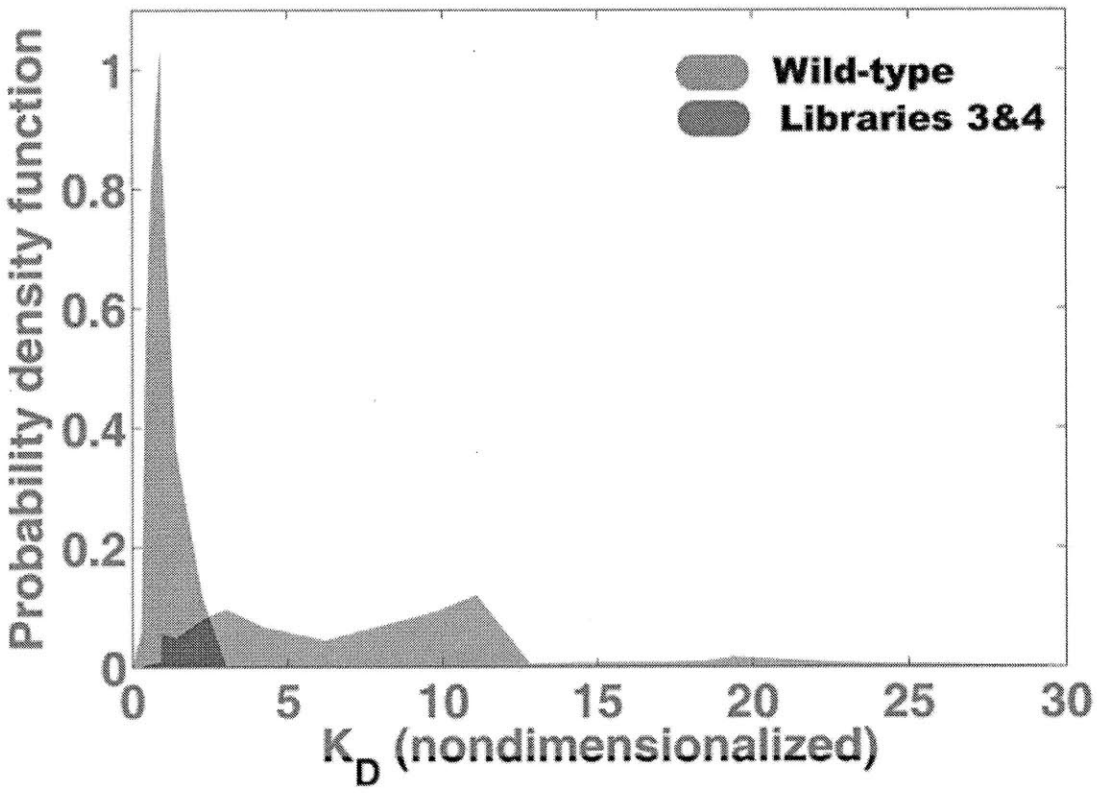


Figure A3.7. The probability density functions of K_D. The K_D distributions were inferred from the data in Figure A3.5 using equations [4.1] and [4.2].

Appendix 3.3 References

Gari E., L. Piedrafita, M. Aldea, E. Herrero, *Yeast* **13**, 837 (1997).

Sizemore C., A, Wissmann, U. Gulland, W. Hillen, *Nucl. Acids Res* **18**, 2875 (1990)

Appendix 4 Abbreviation list

chIP	chromatin immunoprecipitation
FACS	fluorescence activated cell sorting

FRAP	fluorescence recovery after photobleaching
FISH	fluorescence in situ hybridization
GTFs	general transcription factors
HAT	histone acetyltransferase
MLE	maximum likelihood estimation
NER	nucleotide excision DNA repair
NLS	nuclear localization signal
PIC	preinitiation complex
PMF	probability mass function
PSAM	Position specific affinity matrix
RSC	remodels the structure of chromatin
RNA Pol II	RNA polymerase II
SAGA	Spt1-Ada-Gcn5-acetyltransferase
SV40	Simian virus 40 large T-antigen
TAF	TATA box binding protein associated factor
TBP	TATA box binding protein
tetO	tet-operator
TF	transcription factor
TFBS	transcription factor binding site
TSS	transcriptional start site
tTA	tet-transactivator
VP16	viral protein 16 from herpes simplex 1