2014

Quantitative studies of mycobacterial single cell dynamics using a synthetic gene oscillator

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

AND

COLLEGE OF ENGINEERING

Dissertation

QUANTITATIVE STUDIES OF MYCOBACTERIAL SINGLE CELL DYNAMICS USING A SYNTHETIC GENE OSCILLATOR

by

WEN-HAN YU

B.S., National Taiwan University, 2002
M.S., National Taiwan University, 2004

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2014

Approved by

First Reader

_____

James E. Galagan, PhD
Associate Professor of Biomedical Engineering and Microbiology

Second Reader

_____

Martin Steffen, PhD
Assistant Professor of Biomedical Engineering

Third Reader

_____

Ahmad S. Khalil, PhD
Assistant Professor of Biomedical Engineering

QUANTITATIVE STUDIES OF MYCOBACTERIAL SINGLE CELL DYNAMICS USING A SYNTHETIC GENE OSCILLATOR

(Order No.)

WEN-HAN YU

Boston University Graduate School of Arts and Sciences

and College of Engineering, 2014

Major Professor: James E. Galagan, Assistant Professor of Bioengineering

ABSTRACT

The gene regulatory interaction network of *Mycobacterium tuberculosis* (Mtb) has been mostly reconstructed, resulting in a huge step towards revealing complicated interconnections between the network and cellular responses. Nevertheless, the temporal dynamics of gene expression in the regulatory network still remain unclear. Such information importantly provides further understandings in cellular behaviors underlying the system in the presence of external stresses such as those from the host immune system and drug treatments. Here I applied an engineering-based approach to the new context of studying the Mtb temporal dynamical system by adding a synthetic perturbation system. To do so, I developed the first synthetic gene oscillator functioning as a circadian clock in mycobacteria based on a transcriptional control circuit. The circuit encodes differentially-activated positive and negative feedback regulators, which enable autonomous, self-sustained oscillatory gene expression with an average period of 4.5 hours. I modeled robustness and tunability of oscillation period and amplitude computationally and verified them experimentally. I also used this oscillatory circuit transcriptionally to regulate one of the transcription factors (KstR), and profiled *in vivo* temporal dynamics of KstR and KstR-regulated gene targets by dual-color fluorescence intensities in single *Mycobacterium*

smegmatics cells. I observed bifunctionality (activation and repression) of KstR-regulated gene expression, and different gene targets exhibited differential, characteristic response times. This constituted a proof-of-concept that one can utilize a genetic oscillator to characterize dynamical regulation of transcription factors underlying a complex regulatory system. Additionally, it presents a potential new strategy for using dynamical system modeling to guide genetic engineering of mycobacteria for future therapeutic applications.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

aTC	Anhydrotetracycline
BCG	Bacillus Calmette–Guérin
Cap	ε-Caprolactam
ChIP	Chromatin Immunoprecipitaton Followed by Sequencing
ECFP	Enhanced cyan fluorescence protein
Ecoli	*Escherichia coli*
LTBI	Latent TB Infection
MCMC	Markov Chain Monte Carlo
M. smegmatics	*Mycobacterium smegmatics*
MTB	*Mycobacterium tuberculosis*
OD	Optical density
ODE	Ordinary differential equation
PDMS	Polydimethylsiloxane
RBS	Ribosomal binding site
RFP	Red fluorescence protein
SVMs	Support vector machines
TB	Tuberculosis
YFP	Yellow fluorescence protein

Chapter 1. Introduction

1.1 Mycobacterium tuberculosis (MTB) and latent TB Infection

Tuberculosis (TB) is a contagious and airborne disease caused by infection of Mycobacterium tuberculosis (Mtb). The global burden of TB remains enormous, with eight million new cases of active TB and two million deaths per year (WHO, 2012). In addition, it is estimated that one-third of the world's population harbors chronic, asymptomatic latent TB infection (LTBI). Latently infected individuals exhibit reduced responses to drug treatment, and represent a main reservoir of disease reactivation (Young et al, 2009). Mtb has developed a successful strategy in the coevolutionary history with human (Comas et al, 2013) to reside in the human host for decades in a dormant phenotype, withstanding host immune responses and other stresses including antibacterial drugs (Russell, 2007). The mechanisms underlying Mtb's ability to reside in the human host in a dormant phenotype have been the subject of intense scrutiny, but remain largely unknown. Decoding the genetic mechanisms underlying Mtb's dormancy adaptation could enable new insights and opportunities for novel drugs targeting LTBI.

1.2 Mtb gene regulatory network construction with hypoxia adaptation

We recently reconstructed a model of Mtb's gene regulatory network using ChIP-Seq based methods (Galagan et al, 2013). Additionally, we integrated this framework with system-wide -omic profiling under conditions of hypoxia and reareation (Galagan et al, 2013). Hypoxia is a key host-induced stress during Mtb infection and has been hypothesized to induce the Mtb dormancy adaptation response (Rustad et al, 2009). Several potential master regulators of hypoxia adaptation and their interconnections to lipid metabolism during dormancy have been found (Galagan et al, 2013; Kendall et al, 2007; Park et al, 2003; Rustad et al, 2008; Singh et al, 2009;

Walters et al, 2006). Chip-Seq based methods combined with cell-population-averaged -omics profiling revealed several key insights about the regulatory network functionality, but these methods only provide information about the steady-state interactions of the network. Thus, temporal dynamics of gene regulatory interactions in the Mtb network remain poorly defined.

1.3 Importance of temporal dynamics

In this report we illustrate the importance of discovering network dynamics. The regulatory network can be decomposed into many network motifs, each of which carries out specific information-processing functions (Alon, 2007). Importantly, these functions are governed not only by the network topology, but also by the kinetic parameters for enzymatic reactions, as well as single-cell gene expression noise. Time-dependent (i.e. dynamic) data provide information connecting previous and current cellular states, and this data is necessary to constrain the parameters of the model and to validate gene interactions in ways that static-state data cannot (Bennett and Hasty, 2009; Locke and Elowitz, 2009). Furthermore, isogenic cells responding to identical environmental stresses simultaneously exhibit stochastic fluctuations in gene expression dynamics, and which lead to cell-to-cell phenotypic variation (Cai et al, 2008; Eldar and Elowitz, 2010; Locke et al, 2011) and/or determine cell fate permanently (Maamar *et al*, 2007; Suel *et al*, 2006; Suel *et al*, 2007); to capture this information, single-cell temporal measurements are necessary.

1.4 Reverse engineering a dynamical system

To reverse-engineer a dynamical system, a well-defined and programmable perturbation is frequently applied to drive the system, and the influence of this perturbation on the system output is measured. In this way, observed input-output data can be used to construct a robust mathematical model of the system (Ljung, 2010; Mettetal et al, 2008). In contrast to this systems

identification methodology, conventional gene regulation studies implement step-like inputs, e.g. gene knockout or overexpression data. This data is not always informative because cells may respond to perturbations transiently and quickly return to steady-state. To better understand design principle of gene circuits in vivo, an engineering approach was suggested that provides periodical input to the network, measures its output, and describes the dynamics by applying a model (Lipan, 2008; Lipan and Wong, 2005; Ljung, 1999). One common model is to use an oscillatory input, which may be generated from cell cycle (Lin et al, 2012), circadian rhythm (Izumo et al, 2006; Konopka and Rooman, 2010), a oscillation gene circuit responded to stresses (Geva-Zatorsky et al, 2010) or an external periodic stimulus (Mettetal et al, 2008; Taylor et al, 2009). The advantages of using oscillatory signals is that oscillations are relatively insensitive to noise when processed by mathematical algorithms such as cross-correlation and Fourier analysis, and innately encode more quantitative parameters (e.g. intensity, period, and slope) than step functions. Additionally, the frequency encrypted in the system output is modulated solely by the input frequency, without influence of internal parameters in the biological system, again in contrast to step inputs (Lipan et al, 2005). Consequently, oscillatory signals provide a unbiased way to validate network interactions (Konopka et al, 2010).

1.5 The architecture of oscillator

A standalone synthetic gene oscillator has been successfully constructed in E.coli (Elowitz and Leibler, 2000; Mondragon-Palomino et al, 2011; Stricker et al, 2008) and mammalian cell (Tigges et al, 2009) that can self-sustain for several cell generations and exhibits tunability of period and amplitude, mediated by the concentration of inducer chemicals. We applied the design principles of a previously-reported gene circuit (Stricker et al, 2008), namely a dual-positive-negative-feedback oscillator, to create an independent oscillatory input generator in

vivo in Mycobacteria. The programmable oscillator was coupled to the native Mycobacterial regulatory network. As results, the output exclusively responded to the synthetic oscillator via network connections. On the other hand, a periodic input can be possibly generated by external periodic stimulus (Taylor et al, 2009). But the external stimulus may simultaneously initialize other unexpected inputs, and consequently the challenges to decode the mechanisms encrypted in signal information rise enormously.

1.6 Project overview

In this report, building on our knowledge of the Mtb regulatory network via a systems biology approach, we reverse-engineered the temporal dynamics of gene regulation interactions using a synthetic-biology approach. Specifically, we constructed an oscillator gene circuits in a bottom-up fashion to modulate expression of well-characterized regulators in Mtb regulatory network. Time-dependent dynamic input-output data were measured at single-cell resolution. This proof of concept demonstrates a new idea to characterize dynamic regulations of transcription factors (TFs) and target genes underlying a complex regulatory network, and also provides both a detailed mathematical model and estimated parameters useful for Mtb research community quantitatively modeling gene regulatory network in the future.

Chapter 2. A synthetic genetic oscillator constructed in mycobacteria with robustness and Tunability

2.1 Introduction

To build the oscillator system in Mycobacteria, as compared to other model organisms such as *E.coli* and Yeast, is highly challenging in many aspects, including slow growing condition of Mycobacteria as well as the lack options for inducible gene expression circuits. In the design concept of the oscillator, we implemented a previously modeled network architecture composing of linked positive and negative feedback loops (Hasty *et al*, 2002; Smolen *et al*, 1998). This model was shown to be robust and highly tunable in terms of frequency and amplitude of the oscillation cycle (Stricker *et al*, 2008). To create a feasible synthetic gene circuit with the properties of dual-feedback loops in Mycobacteria, two well-characterized inducible gene expression systems, including Nitrile and TetR system as positive and negative feedback loops respectively, were implemented in the gene circuit design. Regulation of NitR to *nit*A promoter behaved as a bistable switch and its regulatory activity was the function of inducer ε-Caprolactam concentration (Pandey *et al*, 2009). TetR regulation system where TetR dimer binds to *tetO* promoter to repress the gene expression was also dependent to inducer anhydrotetracycline (aTc) concentration (Ehrt *et al*, 2005). In order to build a dual-feedback-loops framework, we engineered a chimera promoter $P_{nit/tet}$ that *nit*A fused with dual-*tetO* sequences and which was recognized by both NitR and TetR. This hybrid promoter module located at upstream of both NitR and TetR gene as well as the reporter gene encoding red fluorescence protein (Figure 5a).

2.2 Methods

2.2.1 Build the synthetic gene oscillator

For oscillator gene circuit cloning, we started with constructing three operon cassettes for *nit*R, *tet*R and red fluorescence protein *rfp* gene, respectively. The beginning of each contained a *nit*A/*Tet*O chimera promoter in a sequential order of *nit*A – -35 – *tet*O– -10 – *tet*O and followed by a ribosomal binding site (RBS), and in the end of structural gene, ssrA degradation tag followed by a transcriptional terminator sequence were added. Three cassettes were then assembled on a pMV261 plasmid, a shuttle vector between *Ecoli* and mycobacteria (Stover *et al*, 1991). The recombinant plasmids were transformed to *Ecoli* and analyzed by sequencing. The extracted plasmids (pWOsc1) from *Ecoli* were re-transformed to *M.Smegmatics* WT.

2.2.2 Microfluidics device fabrication

Devices are constructed of a patterned transparent polymer, polydimethylsiloxane (PDMS), formed on a silicon-based mold following well-established fabrication techniques (Sia and Whitesides, 2003) and bonded to a glass coverslip. The device contains inlet and outlet holes that are punched during fabrication and allow for the delivery and drainage of media via polyethylene tubing. The architecture of the device is shown in Figure 1. Cells are trapped in a region of the channel (“trap region”) that has a 0.9-micron channel height, in order to allow a monolayer of cells to be immobilized and monitored over long time periods. Media is flowed continuously through this region, accessed via a “flow channel” of 3.0 microns. Fluid delivery is controlled by means of a programmable syringe pump or by pressure-controlled air flow. To avoid collapse of the trap region, a dense grid of PDMS support pillars are incorporated into the device architecture. The entire device and microscopy apparatus is housed in a temperature-controlled chamber that is kept at 37 ºC for the duration of each experiment.

2.2.3 Cells preparation and time-lapse fluorescence microscopy setup

Mycobacteria cells (*M.smegmatics*) cultured in flask overnight (OD600 of approximately 0.4) were diluted and filtered through 5 μm filter (Millipore) to break up cell clumps and avoid device clogging. Appropriate inducers were added to the filtered cell culture before loading this suspension into a syringe. Device channels were flushed with fresh inducer-rich medium prior to cell loading via a syringe that was connected to the device inlet via polyethylene tubing. Cells were injected from a syringe connected to the device via polyethylene tubing connected to the fluid outlet; pressure was applied by manually depressing the syringe plunger. Once cell loading was confirmed visually, the cell-injection syringe and tubing were disconnected from the device. Slow, steady fresh media flow during the experiment was maintained via a programmable syringe pump or by connecting the media tubing to an air pressure manifold. This media constantly supplied the cells with fresh nutrients but was gentle enough to avoid displacement or dislodging of the cells during the experiment.

The microscope system (Nikon) was equipped with automatic focus software, motorized stage and light shutters, and an image acquisition system (NIS-Elements) that enabled time-lapse experiments of >48 hours in duration and continuous monitoring of results. The hardware and software setup followed the instruction of Nikon automatic microscopy system and NIS-elements Imaging software. Before taking time-series phase contrast and fluorescence images, the device was placed on the stage which was surrounded by a temperature-controlled enclosure (Young *et al*, 2012) for 3 hours to reach external equilibrium in term of temperature and flow pressure. A visual field was selected based on the cells in the field that were confirmed to grow during this three hour period. Cells grew exponentially to fill the visual field over an experimental duration of 24~36 hours while the images were periodically acquired every 12.5 min in phase contrast channel and 25 min in fluorescence channel. The 25-minute fluorescent acquisition frequency

was determined to be a sufficient sampling rate that did not lead to significant photobleaching of the unstable fluorescent proteins during the course of the experiment. During image acquisition, focus was maintained by a contrast-based autofocus tracking system (Nikon). At the end of experiments, the multiple-channel images including phase-contrast and varied fluorescence were saved into a single .nd2 file for image processing.

2.2.4 Image processing and cell tracking

There are several steps for image processing including extracting individual channel images at each time-point, aligning the images, and image background subtraction. We accomplished each of these tasks using Miji, a java-based package, to run image processing package- Fiji (ImageJ development version) within the MATLAB environment. Miji uses Matlab as user interface and processing core and sends results to Fiji. For image alignment, the position of the visual field may be slightly drifted in a long period of time of an experiment due to background vibration from environment, and lack of position correction might raise the difficulty for subsequent cell tracking processes. Therefore, the Fiji plugin -'Register Virtual Stack Slices' were implemented to align the phase-contrast images through all time-series data and to store transformation, and followed by calling the second plugin - 'Transform Virtual Stack Slices' to correct the relative position of other fluorescence images at the same time point from transformation information. Next background subtraction tried to homogenize uneven illuminated background which improves accurate quantitation for fluorescence signals. After these processes, individual channel images were saved for next cell tracing analysis.

Movies that integrated phase-contrast and multiple fluorescence channel images by the Fiji plugin-Image5D were produced and the input-output fluorescence colors were overlaid in the

movie for better visualization. It captured temporal dynamics of input RFP signals and output YFP signals at time-series points and coupled with cell growth and division events.

For image segmentation and cell tracing, we implemented MicrobeTracker- a Matlab-based software package (Sliusarenko *et al*, 2011) in our image processing pipeline. MicrobeTracker provides a set of tools to measure the properties of spatio-temporal organization. It can precisely outline the rode-shaped bacteria cells based on phase-contrast images and track the cells in time lapses over multiple generations. With the power of tracing the cells and its progeny, cell lineages over the generations during experimental period of time can be revealed. In addition, the software superimposes the identified cell outlines on fluorescence images at each time point in which the fluorescently labeled objects can be quantified at subpixel resolution. Furthermore, the program provides an interface in Matlab environment where the users are able to visualize the processes and manually identify and modify cell outlines.

However, despite the fact that mycobacterial cells are classified as rod-shaped cells but they exhibit unique morphology distinguishing from the model rod-shaped organisms such as *E. coli* and *B. subtilis*. Mycobacteria grow and divide in the fashion of asymmetry at a wide range of cell size and snap into a characteristic V-shape upon division (Aldridge *et al*, 2012; Hett and Rubin, 2008; Joyce *et al*, 2012). In addition, mycobacteria tend to aggregate to densely packed cells because of high lipid content in the cell wall and causing a hydrophobic interaction (Borrego *et al*, 2000). Furthermore, from empirically analyzing mycobacterial images, inhomogeneous phase-contrast intensities within a single cell were frequently observed due to perhaps uneven surface of the cells. All of unique phenotypes raised the poor performance of MicrobeTracker for analyzing mycobacterial cells automatically, especially the program often failed to identify when and where the dividing points of the cells happen.

In order to automatically and robustly trace mycobacterial cells and improve the accuracy of division events identification, we have modified MicrobeTracker by adding a supervised learning function (support vector machines, SVMs) in the processes of cell contour defining. MicrobeTracker for cell splitting detection was used to scan the “phase-contrast intensity” valleys along the profile of each cell, and the potential dividing points satisfied the ratio of decreased intensity by the valley depth versus mean of maximum intensities in both cell area split by the point. Such algorithm often failed due to poor thresholding selection (contrasting objects versus uniform background) and uneven surface of the cells. Therefore, a manually derived set of cell contours was provided to train the SVMs on a set of visual features including the changes on 1-dimensional intensity profile and the shape of the cells on 2-dimensional coordinates. Then after new cells feeding in, SVMs based on the training model classified each geometric point along the profile (dividing or not dividing). With SVMs, the program greatly diminished the time for manually mycobacterial cell contour correction. SVMs ran under Matlab environment as functions in MicrobeTracker package. Finally, to double check whether the cell tracing has been suggested correctly by the program, we additionally analyzed the sizes of daughter cells after dividing. The sum of two daughters’ sizes should be similar to the parent cell right before dividing. The size difference ratio should be less than the empirical threshold value of 3.0.

2.2.5 Signal processing and phenotypic characteristics extraction

The outputs of MicrobeTracker included cell outline coordinates on the images, cell size, and fluorescence intensities of each cell in time-series images as well as cell lineages over the generations. The fluorescence trajectories of individual cells in the experiment were first normalized by cell size and noises-filtered by selecting a smoothing spline fit algorithm in Matlab. Each cell input-output signal trajectories were plotted individually. Additionally we

analyzed cross correlation function between input and output among all cells which were identified under the objective field. By plotting mean of cross correlation function over all single cells on the time domain, we can identify the time delay of output responding to input as well as the sign of the input regulation which was indicated by the sign of correlation coefficient (Figure 12b).

With quantitative data converted and extracted from a set of live-cell time-series images, there are many more mycobacterial physiological characteristics which are challenged in the past and can be discovered from our image analysis. For instance, a fundamental question can be asked as whether size or time-dependent manner controls mycobacteria division. By tracing the cells over the generations, the relationship between cell size and doubling time may provide insight to this question (Figure 15). Additionally, it is still unclear how cellular components distributed into two daughter cells during division and which mainly causes cell heterogeneity in term of cell regulatory responses. With integration of single-cell signal trajectories and cell lineage information, it provides great insight about how the signals and noises propagate from progenitor through its progeny.

The whole processing pipeline from raw image file to final signal analysis was performed using custom-written Matlab scripts with MicorbeTracker (downloaded from http://microbetracker.org/) and Fiji (downloaded from http://fiji.sc/Fiji) integrated in Matlab by Fiji plugin-Miji (http://fiji.sc/Miji).

2.3 Results

2.3.1 Build the oscillator gene circuit

We successfully developed the first synthetic circadian clock driving oscillatory gene expression in Mycobacteria (Figure 5a), which has been known a challenge for Mycobacteria re-

engineering, due to a lack of inducible gene circuit options as well as the physical parameters for computational model prediction. In the design concept of the oscillator, we implemented a previously modeled network architecture composing of linked positive and negative feedback loops (Hasty et al, 2002; Smolen et al, 1998). This model was experimentally and mathematically demonstrated to be robust and highly tunable in terms of frequency and amplitude of the oscillation cycle (Stricker et al, 2008). To create the gene circuits with the properties of dual-feedback gene circuits, we engineered a chimeric promoter region (Pnit/tet) that contained a activation operator site from nitA promoter located at upstream of -35 bps region as RNA polymerase binding site, and two repression operator sites from tetO promoter located at the franking regions of -10 bps region as the other RNA polymerase binding site. The binding affinity of NitR to the operator nitA site was modulated by the concentration of ε-Caprolactam (Cap). The positive cooperative function between the binding affinity and Cap has been shown in a previous report and was parameterized for a computational model (Pandey et al, 2009) (Ch 3.2.1.2; Figure 7). Similarly, TetR binding affinity was modulated by the concentration of anhydrotetracycline (aTC), in that aTC bound TetR and caused reduction of the binding affinity to dual-operator tetO sites (Ehrt et al, 2005) (Ch 3.2.1.2; Figure 8). The chimeric promoter Pnit/tet kept similar kinetic characteristics of regulations in response of Cap concentrations (Figure 2). Additionally, the sequential positions of the operator sites (nitA – -35 – tetO – -10 – tetO) in Pnit/tet at upstream of transcription starting site leaded to a logic combinational regulation between NitR and TetR where target gene expression activated only when NitR exclusively bound to the region, whereas the expression repressed in presence of TetR. This regulation was demonstrated by measuring fluorescence reporter expression under the control of Pnit/tet (Figure 3) and implemented in the computational model construction (Eq. 3.2.1-1 to 3). NitR, TetR and the fluorescence reporter

protein were all under the control of Pnit/tet. As results, the designed circuits interconnected between the activator and repressor in the dual-feedback loops (Figure 5a).

2.3.2 The mechanism of oscillation

Here we describe how oscillator worked from mechanistic aspect (Figure 5b). There are four states in the oscillator, representing the combinatorial bindings of NitR and TetR. At the initial state, in the absence of both aTC and Cap, small amount of both TetR and NitR are existed in the cells, but only TetR is able to bind to Pnit/tet, because NitR keeps inactive without Cap. When both inducers added to the cell medium, aTC binding to TetR triggers TetR conformational changes and reduces the binding affinities (Km), which leads to TetR disengaging from Pnit/tet (Second state). Oppositely NitR binding affinity increases via Cap binding to NitR. At this stage, both NitR and TetR start to slowly express, but Pnit/tet is most likely bound by NitR because of the changes of Km, and a delay time for TetR forming an active dimer. As a result, it turns on the positive feedback loop and leads to the third stage. Meanwhile, TetR dimer was also synthesized under Pnit/tet control. With a time-delayed accumulation of active TetR dimer, TetR dimer also binds to Pnit/tet and activates the negative feedback loop (Forth state). After that, the synthesis of NitR and TetR shuts down and the degradation takes over. Finally the numbers of the regulators dramatically decline and the system is back to the second state. Consequently, timely differential responses of positive and negative feedback loops can drive oscillatory behavior (Figure 5c, d).

2.3.3 The selection of ssrA degradation tag and tunability of oscillation

Furthermore, protein degradation coupled with protein synthesis contributed to shaping temporal activator and repressor expression profiles. All three genes were directly linked with carboxyl-terminal ssrA tags which were comprised of 11 amino acids (AANDENYALAA) and recognized by ClpXP and ClpAP for protein degradation (Karzai et al, 2000). Based on the

previously reported theoretical design (Hasty et al, 2002), oscillation was favored when activator and repressor had differential protein degradation rates. Therefore we engineered three tags with different degradation rates (fast, medium and slow) by mutating the residues on 11-amino acids (Chang et al, 2008; Flynn et al, 2001), and tested all combinations of activator and repressor linked to different tags in order to search optimal oscillatory condition (Figure 4). From 9 combinatorial tests, activator linked to medium-speed degradation tag and repressor linked to fast/slow tag exhibited potential robust oscillatory behavior. Others showed damped oscillatory signals and approached to either high or low steady state within 10 hrs measurement. We selected a strain with the activator containing a medium tag and the repressor containing a fast tag (pWOsc4, Table 1), for the following experiments, and coupled it with mycobacteria regulatory network. To investigate tunability of the genetic oscillator, we titrated the inducing concentration of both inducers (Cap and aTC) and measured the changes of amplitude and period of oscillatory fluorescent signals of M. Smegmatis cells cultured in the early exponential phase (Figure 10a, b; black triangle with error bars). When Cap concentration was raised at a fixed value of 20 ng/ml aTC, period of the oscillation decreased, whereas the amplitude increased. Period and amplitude approached steady state at Cap concentrations of >20 mM. Next, Cap concentration was fixed at 14 mM and varying aTC concentrations were tested, which also allowed for tuning both period and amplitude (Figure 10c, d; black triangle with error bars). The inducer titration curves indicate the robustness of the oscillator in mycobacterial cells with a wide range of both inducer concentrations.

2.3.4 Oscillation at single cell resolution

In order to examine oscillatory behavior in individual cells, we have developed and used a microfluidics device that allows a monolayer of mycobacteria cells to be immobilized and to

grow for several generations, with a constant flow of nutrient medium into the device (Figure 1). All single cells in the field of view of a 100X-magnification microscope were monitored and traced through the time-course experiments (Figure 13). Individual cell positions identified from phase contrast images were superimposed on fluorescence images, and phenotypic characteristics of the individuals such as gene fluorescence expression, cell size, doubling time and cell lineage were extracted quantitatively (Figure 14). The individual cells exhibited averaged five cycles of fluorescence oscillations in 24 hours of measurements, with inducer concentrations of 25 ng/ml aTC and 14 mM Cap (Figure 5d, Figure 6). Interestingly, comparing to the published bacterial oscillator in Ecoli (Stricker et al, 2008), the period of the oscillation in our model (4.5 hrs in average) was much longer than in Ecoli (0.5 hr). The different growth rates between E.coli and Mycobacteria was suggested to be one of the major factors to contribute the length of oscillation cycle (Stricker et al, 2008). The oscillatory phases between daughter cells showed variability, suggesting that cellular internal noises, e.g. asymmetric division in mycobacteria (Aldridge et al, 2012), may also influence oscillatory phase and result in desynchronized oscillations among individual cells after several cell generations.

Chapter 3. System identification and dynamical system analysis

3.1 Introduction

To gain insight into the temporal dynamics of each molecular component in the oscillator gene circuits and its connected and driven regulatory network in Mycobacteria where we constructed to generate nature of oscillation, the complete biochemical reaction models for the synthetic gene circuits are constructed. The models incorporate the effects of biochemical interactions including the binding of inducers to transcription factor, the transcription factor binding to the promoters of the target genes, transcription of mRNA, translation of polypeptides and dimerization, degradation of mRNA and protein and the actual gene copy number. Our models assume that the molecular components are homogonously distributed in the cells as well as after cells duplicated into two siblings. We used SimBiology toolbox in Matlab to construct complete ODE kinetic reaction model and simulate the system dynamics for individual components.

3.2 Methods

3.2.1 Ordinary differential equation (ODE) of kinetic reactions in oscillator model

3.2.1.1 Transcription factor binding to promoter

The promoter regions of NitR, TetR and reporter gene that form the core of the gene circuit are all composed of an identical hybrid promoter that contain one NitR binding motif and followed by two TetR binding motifs.

$$P_{0,i}^{j} + a \underset{k_{-a}}{\overset{k_a}{\rightleftharpoons}} P_{1,i}^{j} \qquad i \in \{0,1,2\};\ j \in \{a,r,fp\}$$ Eq. 3.2.1-1

$$P_{s,0}^{j} + r_2 \underset{k_{-r}}{\overset{2k_r}{\rightleftharpoons}} P_{s,1}^{j} \quad s \in \{0,1\};\ j \in \{a,r,fp\}$$ Eq. 3.2.1-2

$$P_{s,1}^{j} + r_2 \underset{2k_{-r}}{\overset{k_r}{\rightleftharpoons}} P_{s,2}^{j} \quad s \in \{0,1\};\ j \in \{a,r,fp\}$$ Eq. 3.2.1-3

where a and r_2 are NitR monomer and TetR dimer. P is the hybrid promoter and i and s indicates the amount of TetR dimer and NitR monomer binding to promoter. j presents the gene (a=NitR; r=TetR; f=reporter gene) where the promoter is bound. Eq. 3.2.1-1, 2, and 3 displayed the regulator association and dissociation reactions on the target promoter.

3.2.1.2 The binding rate of transcription factor in presence of inducers

The forward NitR and TetR binding rates (*ka* and *kr*) can be written as the function of Caprolactam and aTC concentrations respectively in the Hill equations. Eq. 3.2.1-4 indicates the independent binding of TetR which is not influenced by the presence of NitR. Eq. 3.2.1-5 indicates NitR binding when it exists alone.

$$k_r = k_{-r} \cdot \left(\left(B_r^{\max} - B_r^{\min} \right) \cdot \frac{1}{\left(1 + \left(\frac{aTC}{km_r} \right)^{h_r} \right)} + B_r^{\min} \right)$$ Eq. 3.2.1-4

$$k_a = k_{-a} \cdot \left(\left(B_a^{\max} - B_a^{\min} \right) \cdot \frac{Cap^{h_a}}{\left(km_a^{h_a} + Cap^{h_a} \right)} + B_a^{\min} \right)$$ Eq. 3.2.1-5

where $B_{a/r}$ present maximum and minimum binding affinity of NitR monomer and TetR dimer to the promoter. $h_{a/r}$ present the hill coefficients and $km_{a/r}$ are the concentrations of Cap and aTC for producing half of binding affinity.

3.2.1.3 Synthesis of mRNA and protein

The following equations describe mRNA transcription, protein translation, folding and dimerization. In the absence of TetR and NitR binding to promoter, the target gene shows basic transcription (*Ts*) and once NitR binding without the presence of TetR, the transcription rate increase α fold change.

$$P_{0,0}^{j} \xrightarrow{Ts} P_{0,0}^{j} + m_j \qquad j \in \{a,r,fp\}$$ Eq. 3.2.1-6

$$P_{1,0}^{j} \xrightarrow{\alpha \cdot Ts} P_{1,0}^{j} + m_j \qquad j \in \{a,r,fp\}$$ Eq. 3.2.1-7

$$m_a \xrightarrow{Tl} m_a + a$$ Eq. 3.2.1-8

$$m_r \xrightarrow{Tl} m_a + r$$ Eq. 3.2.1-9

$$m_{fp} \xrightarrow{Tl} m_{fp} + fp$$ Eq. 3.2.1-10

$$r + r \underset{k_{-d}}{\overset{k_d}{\rightleftarrows}} r_2$$ Eq. 3.2.1-11

where *Tl* is translation rate and k_d/k_{-d} are protein dimerization and de-dimerization rate. m indicates message RNA of the regulators and the reporter. Eq. 3.2.1-6 and 2 displayed basic and active transcription reaction; Eq. 3.2.1-8, 4, and 5 showed translation reaction and Eq. 3.2.1-11 as dimerization reaction.

3.2.1.4 Dilution and degradation of mRNA and protein

Total amount of mRNAs and proteins are subjected to dilution rate due to cell duplication and degradation rate by RNase and protease in the cell. Protein degradation rate of the targets are determined by ssrA tag.

$$m_j \xrightarrow{d+du} \varnothing \qquad j \in \{a,r,fp\}$$ Eq. 3.2.1-12

$$a \xrightarrow{\delta \cdot f(ssrA)+du} \varnothing$$ Eq. 3.2.1-13

$$r \xrightarrow{\beta \cdot f(ssrA)+du} \varnothing$$ Eq. 3.2.1-14

$$r_2 \xrightarrow{\beta \cdot f(ssrA)+du} \varnothing$$ Eq. 3.2.1-15

$$fp \xrightarrow{\kappa \cdot f(ssrA)+du} \varnothing$$ Eq. 3.2.1-16

$$\text{Where} \quad f(ssrA) = \frac{\gamma}{Kp + ssrA}$$

Eq. 3.2.1-12 indicated mRNA decreased reaction and Eq. 3.2.1-13 to 5 presented protein decreased reactions. For protein degradation, we assume that the number of proteases is limited compared to the free proteins with degradation tag in the cell. Therefore, *f* here follows Michaelis-Menten kinetics. γ indicates the maximum degradation rate and *Kp* is substrate concentration at which the reaction reach half of reaction rate, and ssrA shows total amount of proteins with ssrA tags. δ, β and κ present the fold change values after ssrA tag are modified to different degradation rates. *du* indicates cell dilution rate.

3.2.2 Data fitting and parameters estimation

The whole parameters in model were showed in the table (Table 2), and can be categorized into three groups in term of the way to acquire. First group of the parameters were able to be found on the literature and which can directly apply in our model prediction. For second group of parameters, we decomposed the model into individual simple reactions and

estimated the parameters in those reactions by using experimental data in the assumption of steady-state condition. Third group of parameters relied on a statistical approach - Bayesian MCMC algorithm to estimate the posterior distributions of those parameters by searching the parameters space.

3.2.2.1 NitR and TetR data fitting to Hill functions

Two papers (Ehrt, 2005 and Pandey, 2009) individually studied NitR positive feedback and TetR negative feedback inducible system in Mycobacteria. Both showed that the fluorescence intensities of the reporters regulated by NitR or TetR are dependent to inducer titrations after inducers were added for overnight and the reactions reached equilibrium. Therefore, the data will be used to estimate the parameters in the Hill function that describes the cooperative binding of the inducers to transcription factor. Firstly we assume that the fluorescence intensities of the reporters can be represented the transcription factor binding to promoter rate according to the following mathematical statement.

To give a proof-of-principle, we take NitR inducible system as an example. Five kinetic reactions describe NitR activating fluorescence reporter gene expression (Eq. 3.2.2-1, NitR binding; Eq. 3.2.2-2, transcription; Eq. 3.2.2-3, mRNA degradation; Eq. 3.2.2-4, translation; Eq. 3.2.2-5, protein degradation).

$$P_0^{fp} + a \underset{k_{-a}}{\overset{k_a}{\rightleftarrows}} P_1^{fp}$$ Eq. 3.2.2-1

$$P_1^{fp} \xrightarrow{\beta} P_1^{fp} + m_f$$ Eq. 3.2.2-2

$$m_{fp} \xrightarrow{\mu_m} \varnothing$$ Eq. 3.2.2-3

$$m_{fp} \xrightarrow{Tl} m_{fp} + fp$$ Eq. 3.2.2-4

$$fp \xrightarrow{f} \varnothing$$ Eq. 3.2.2-5

where β is active transcription rate and μ_m and μ_p are mRNA and protein degradation plus dilution rate. Those can be included into three ODEs (Eq. 3.2.2-6,7 and 8).

$$\frac{dP_1^{fp}}{dt} = k_a \cdot a \cdot P_0^{fp} - k_{-a} \cdot P_1^{fp}$$ Eq. 3.2.2-6

$$\frac{dm_{fp}}{dt} = \beta \cdot P_1^{fp} - \mu_m \cdot m_{fp}$$ Eq. 3.2.2-7

$$\frac{dfp}{dt} = Tl \cdot m - f \cdot fp$$ Eq. 3.2.2-8

After fluorescence intensity reaches steady state, Eq. 3.2.2-8 should be equal to zero as well as Eq. 3.2.2-6 and Eq. 3.2.2-7 because the transcription factor binding rate and transcription rate are much faster than translation rate. So,

$$P_{1\ st}^{fp} = \frac{k_a}{k_{-a}} \cdot a \cdot P_0^{fp}$$ Eq. 3.2.2-9

$$m_{fp_{st}} = \frac{\beta}{\mu_m} \cdot P_1^{fp}$$ Eq. 3.2.2-10

$$fp_{st} = \frac{Tl}{f} \cdot m_{fp}$$ Eq. 3.2.2-11

Substitute Eq. 3.2.2-9, 10 and 11 into Eq. 3.2.2-12

$$fp_{st} = \frac{Tl}{f} \cdot \frac{\beta}{\mu_m} \cdot \frac{k_a}{k_{-a}} \cdot a \cdot P_0^{fp} \qquad \text{Eq. 3.2.2-12}$$

Recall Eq. 3.2.1-5, we can rewrite to Eq. 3.2.2-13

$$fp_{st} = \frac{Tl}{f} \cdot \frac{\beta}{\mu_m} \cdot a \cdot P_0^{fp} \cdot \left(\left(\mathrm{B}_a^{\max} - \mathrm{B}_a^{\min} \right) \cdot \frac{\mathrm{Cap}^{h_a}}{\left(km_a^{h_a} + \mathrm{Cap}^{h_a} \right)} + \mathrm{B}_a^{\min} \right) \qquad \text{Eq. 3.2.2-13}$$

In the experimental designs of Ehrt 2005 and Pandey 2009, inducer concentration is the only variable titrated among different samples. As results, we approximate *Tl*, β, μ_m and μ_p as constant. The number of NitR binding site in the promoter of reporter gene on the plasmid P_0^f also keeps constant among the cells. In addition, amount of NitR – "a" driven by a constitute promoter reach a constant level after equilibrium state. Therefore,

$$fp_{st} \propto \left(\left(\mathrm{B}_a^{\max} - \mathrm{B}_a^{\min} \right) \cdot \frac{\mathrm{Cap}^{h_a}}{\left(km_a^{h_a} + \mathrm{Cap}^{h_a} \right)} + \mathrm{B}_a^{\min} \right) \qquad \text{Eq. 3.2.2-14}$$

3.2.2.2 NitR binding affinity parameter estimation

According to Pandey 2009, *Mycobacteria Smegmatics* were treated by varied caprolactam concentrations (0-57 mM) and reporter gene expression was assayed after 16 hrs incubation. The caprolactam titration curve is shown in the figure (blue line) (Figure 7), and simulations is presented in red dash line. The parameters from Eq. 3.2.2.1-10 are estimated after curve fitting. B_a^{max} = 1.0235 molecule^{-1}; B_a^{min} = 0.0014 molecule^{-1}; km_a = 18.8mM ; h_a = 3.3 .

3.2.2.3 TetR binding affinity parameter estimation

Ehrt 2005 transformed the plasmid carried TetR system with GFP as reporter gene into M.Smegmatics and treated the cells with varied aTC concentrations (0-500 ng/ml). The fluorescence outputs were measured after 15 hrs incubation. The aTC titration curve is shown in the Figure 8 with blue line from actual experimental data and red dash line from simulation after data fitting. Ehrt 2005 indicated that high concentration of aTC showed to reduce cell growth. Therefore, fluorescence signals have been divided by OD before normalized by maximum of the signals. The parameters are estimated. B_r^{max} = 0.9857 molecule^{-1}; B_r^{min} = 0.0857 molecule^{-1}; km_r = 22.7 ng/ml; h_r = 3.0.

3.2.2.4 Modeling the binding equilibrium constant with combination of Caprolactam and aTC

In order to provide accurate predictions to our gene-circuit designs for oscillator, we applied all parameters estimated from Ch 3.2.1.2 to model the correlation between transcription factor binding equilibrium constant of NitR and TetR ($Kc = {k_a}/{k_{-a}}$ and ${k_r}/{k_{-r}}$) and inducer concentrations. *Kc* represents the net flux between association and dissociation of transcription factor to promoter and which can be modified by the inducer concentration. In our design, each of target genes contains a hybrid promoter that can be regulated by both NitR and TetR. First we show the model where increasing of [aTC] reduces TetR Kc (Figure 9b). It indicates that high [aTC] increase the probability of free promoters unbound by TetR and helps to start gene transcription. Similarly we further modeled NitR *Kc* which is function of [Caprolactam]. From the results in the figure (Figure 9a), the presence of Caprolactam help rise *Kc* which means to increase the probability of hybrid promoter bound by NitR and activate target gene expression.

3.2.2.5 Protein degradation rate parameter estimation

In order to calculate protein degradation rate, we treated the cells containing oscillator gene circuit with high concentration of aTC (100 ng/ml) and Caprolactam (56 mM). In such a condition, TetR repression is eliminated by adding aTC to oversaturate free TetR. Oscillation behavior turns to a step response (Figure 11a). After incubation with inducers overnight (16 hrs), the fluorescence reporter gene expression reached equilibrium. Next the cells were switched into the condition in absence of both inducers which causes NitR dissociation from the promoter and TetR rebinding to the promoter. In such a case, synthesis of the fluorescence reporter protein is assumed to completely shot down. The exponential decay of fluorescence signal represents sum of protein degradation and dilution rate. A simplified ODE describes the synthesis of reporter.

$$\frac{dfp}{dt} = \pi - (f + du) \cdot fp \qquad \text{Eq. 3.2.2-15}$$

where π indicates the synthesis of reporter protein *fp* and *f* shows protein degradation rate and *du* is dilution rate. π is equal to 0 since the synthesis completely stop. Therefore we can solve Eq. 3.2.2-15by

$$fp(t) = fp_{st} \cdot e^{-(f+du)\cdot t}$$

$$\ln(fp(t)) = \ln(fp_{st}) - (f + du) \cdot t \qquad \text{Eq. 3.2.2-16}$$

Basically, since dilution rate *du* can be estimated by cell doubling time (In our case *M.smegmatics* with Δ*kstR* is 4.2 hr (*du*=0.002 min^{-1}) measured from time-lapse microscopy experiment at single-cell resolution), Slow, medium, and fast protein degradation rate can be estimated from our experimental results as f = 0.0047 min^{-1}, 0.011 min^{-1}, 0.034 min^{-1}, respectively.

3.2.2.6 Parameter estimation from Bayesian MCMC algorithm

Bayesian Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm provides an statistical estimation for the posterior parameter distribution given a set of experimental observations and optional prior parameter distributions (Eydgahi *et al*, 2013). According to Bayes' theorem, the conditional probability of free parameters given an observed data set as followed,

$$p(\Phi \mid data) = \frac{p(data \mid \Phi)\, p(\Phi)}{p(data)} \qquad \text{Eq. 3.2.2-17}$$

where Φ indicates the parameter vector in the model and $p(\Phi|data)$ represents the posterior of Φ; $p(data|\Phi)$and $p(\Phi)$ are the likelihood and prior of Φ, respectively. Since the ratio of posterior is only required during MCMC sampling (Eq. 3.2.2-22), we can simplify Eq. 3.2.2-17into Eq. 3.2.2-18

$$p(\Phi \mid data) \propto p(data \mid \Phi)\, p(\Phi) \qquad \text{Eq. 3.2.2-18}$$

Therefore, the value of the log posterior is sum of the log likelihood and log prior as following

$$-\ln\left(posterior(\Phi)\right) = -\ln\left(likelihood(\Phi)\right) - \ln\left(prior(\Phi)\right) \qquad \text{Eq. 3.2.2-19}$$

Then $-ln(\text{likelihood}(\Phi))$ was calculated by the sum of squared deviation between the model and the observed data for a given set of parameters, which was based on the assumption that the measurement errors from the observed data at time *t* followed Gaussian distribution.

$$-\ln\left(likelihood(\Phi)\right) = \sum_t \sum_i \frac{1}{2\sigma_d^2(t)} \left(X_d^i(t) - X_m(t,\Phi)\right)^2 \qquad \text{Eq. 3.2.2-20}$$

where $X_d^i(t)$ displays the observed value at the time point *t* in *i*th of repeated experiments and $X_m(t,\Phi)$ is the value at time *t* from model simulation given a set of parameters Φ. σ_d^2 represents

the variance of observed data set. Next $-ln(\text{prior}(\Phi))$ was calculated according to the assumption where the parameters were independent log-norm random variables.

$$-\ln\left(prior\left(\Phi\right)\right)=\sum_{j}\frac{1}{2\sigma_j^2}\left(\theta_j-\left\langle\theta_j\right\rangle\right)^2 \qquad \text{Eq. 3.2.2-21}$$

where σ_j^2 and $\langle\theta_j\rangle$ are variance and mean of the parameter θ_j. In the end, in order to sample the parameter posterior distribution in the parameter vector space, MCMC walk was implemented and the times of visit at a particular position in the parameter space is proportion to posterior probability. During random walk, the decision of jumping to next position in parameter space was followed by Metropolis–Hastings algorithm,

$$\Phi_{s+1}=\begin{cases}\Phi' & \text{if } Unif\,(0,1)\le\alpha\left(\Phi_s,\Phi'\right)\\ \Phi_s & \text{otherwise}\end{cases}$$

$$\alpha\left(\Phi_s,\Phi'\right)=\min\left(1,\frac{posterior(\Phi')}{posterior(\Phi_s)}\right) \qquad \text{Eq. 3.2.2-22}$$

where Φ_s indicates the current position and Φ' the candidate for next position; Φ_{s+1} the next position.

We downloaded and implemented MCMC toolbox (http://helios.fmi.fi/~lainema/mcmc) (Haario *et al*, 2006) in Matlab. For initial parameter prior setting, we used the parameters available in different organism such as *E.coli* as mean and gave a fixed variance for prior Gaussian distribution. Upper and lower bounds of the parameters also determined according to its mean. In addition, we used a modified approach-DRAM available in the toolbox to improve the efficiency of MCMC algorithm, which was useful in our situations where some of the parameters

were unidentified by the data or were highly correlated. Cost function used sum of least squared errors between single cell normalized fluorescence trajectories and simulation results. In the MCMC walks of > 10^6 steps, all 15 parameters reached convergence by Geweke criteria. The parameters available in different prokaryotes were implemented as mean and a fixed variance was given for its prior Gaussian distribution. We applied the optimal value of the parameters in term of probability into next dynamical system analysis and final model construction.

3.2.3 Stability analysis for dynamic oscillatory system

In this section, we use the quantitative model from section 1 that described the oscillatory system and further simplify the model to only several equations, in order to reduce the complexity of the dynamical system. The kinetic reactions from section 1 can be categorized into two classes, fast and slow, in term of time scale of the reaction rate. The fast reactions show the rate constants of order seconds and can be viewed as in the steady state from the perspective of slow reactions which have the rates of order minutes or hours.

Fast reactions include NitR (Eq. 3.2.3-1, 2, and 3) and TetR (Eq. 3.2.3-4 and 5) binding to promoters and dimerization of TetR (Eq. 3.2.3-6).

$$P_{00}^{j} + a \underset{k_{-a}}{\overset{k_a}{\rightleftarrows}} P_{10}^{j} \qquad j \in \{a, r\} \qquad \text{Eq. 3.2.3-1}$$

$$P_{01}^{j} + a \underset{k_{-a}}{\overset{k_a}{\rightleftarrows}} P_{11}^{j} \qquad \text{Eq. 3.2.3-2}$$

$$P_{02}^{j} + a \underset{k_{-a}}{\overset{k_a}{\rightleftarrows}} P_{12}^{j} \qquad \text{Eq. 3.2.3-3}$$

$$P_{00}^{j} + r_2 \underset{k_{-r}}{\overset{2k_r}{\rightleftarrows}} P_{01}^{j} + r_2 \underset{2k_{-r}}{\overset{k_r}{\rightleftarrows}} P_{02}^{j} \qquad \text{Eq. 3.2.3-4}$$

$$P_{10}^{j} + r_2 \underset{k_{-r}}{\overset{2k_r}{\rightleftarrows}} P_{11}^{j} + r_2 \underset{2k_{-r}}{\overset{k_r}{\rightleftarrows}} P_{12}^{j} \qquad \text{Eq. 3.2.3-5}$$

$$2r \underset{k_{-d}}{\overset{k_d}{\rightleftarrows}} r_2 \qquad \text{Eq. 3.2.3-6}$$

Slow reactions include transcription (Eq. 3.2.3-7 and 8), translation (Eq. 3.2.3-9 and 10), mRNA degradation (Eq. 3.2.3-11), and protein degradation (Eq. 3.2.3-12, 13, and 14)

$$P_{00}^{j} \xrightarrow{Ts} P_{00}^{j} + m_j \qquad j \in \{a, r\}$$ Eq. 3.2.3-7

$$P_{10}^{j} \xrightarrow{\alpha \cdot Ts} P_{10}^{j} + m_j$$ Eq. 3.2.3-8

$$m_a \xrightarrow{Tl} a + m_a$$ Eq. 3.2.3-9

$$m_r \xrightarrow{Tl} r + m_a$$ Eq. 3.2.3-10

$$m_j \xrightarrow{d+du} \varnothing$$ Eq. 3.2.3-11

$$a \xrightarrow{\delta \cdot f + du} \varnothing$$ Eq. 3.2.3-12

$$r \xrightarrow{\beta \cdot f + du} \varnothing$$ Eq. 3.2.3-13

$$r_2 \xrightarrow{\beta \cdot f + du} \varnothing$$ Eq. 3.2.3-14

We can write ODEs to describe the evolution of the concentration of NitR and TetR

TetR

$$\frac{\partial r}{\partial t} = -2 \cdot k_d \cdot r^2 + 2 \cdot k_{\_d} \cdot r_2 + m_r \cdot Tl - \left(\beta \cdot f + du\right) \cdot r$$ Eq. 3.2.3-15

$$\frac{\partial m_r}{\partial t} = P_{00}^{r} \cdot Ts + P_{10}^{r} \cdot \alpha \cdot Ts - \left(d + du\right) \cdot m_r$$ Eq. 3.2.3-16

NitR

$$\frac{\partial a}{\partial t} = m_a \cdot Tl - \left(\delta \cdot f + du\right) \cdot a$$ Eq. 3.2.3-17

$$\frac{\partial m_a}{\partial t} = P_{00}^{a} \cdot Ts + P_{10}^{a} \cdot \alpha \cdot Ts - \left(d + du\right) \cdot m_a$$ Eq. 3.2.3-18

where Eq. 3.2.3-15 and 16 represent the dynamics of TetR protein and mRNA concentration and Eq. 3.2.3-17 and 18 show the dynamics of NitR protein and mRNA concentration. Next we apply the fact that fast reactions are in the steady state at time scale of order minute in the slow reactions. We can eliminate variables as following algebraic expression

$$Ka = \frac{k_a}{k_{-a}} \quad Kr = \frac{k_r}{k_{-r}} \quad Kd = \frac{k_d}{k_{-d}}$$ Eq. 3.2.3-19

$$r_2 = Kd \cdot r^2$$ Eq. 3.2.3-20

$$P_{10}^j = Ka \cdot P_{00}^j \cdot a$$ Eq. 3.2.3-21

$$P_{11}^j = 2 \cdot Ka \cdot Kr \cdot Kd \cdot P_{00}^j \cdot a \cdot r^2$$ Eq. 3.2.3-22

$$P_{12}^j = Ka \cdot Kr^2 \cdot Kd^2 \cdot P_{00}^j \cdot a \cdot r^4$$ Eq. 3.2.3-23

$$P_{01}^j = 2 \cdot Kr \cdot Kd \cdot P_{00}^j \cdot r^2$$ Eq. 3.2.3-24

$$P_{02}^j = Kr^2 \cdot Kd^2 \cdot P_{00}^j \cdot r^4$$ Eq. 3.2.3-25

In addition, total concentration of DNA promoter binding sites should be a constant, so that

$$Pt = P_{00}^j + P_{10}^j + P_{01}^j + P_{11}^j + P_{02}^j + P_{12}^j = P_{00}^j \cdot \left(Ka \cdot a + 1\right)\left(Kr \cdot Kd \cdot r^2 + 1\right)^2$$

Eq. 3.2.3-26

where $P_{10}^j, P_{01}^j, P_{11}^j, P_{02}^j, P_{12}^j$ are substituted by P_{00}^j according to Eq. 3.2.3-19 to 25. Therefore, with substitution of Eq. 3.2.3-19 to 25, we can rewrite Eq. 3.2.3-15, 16, 17 and 18 as following

TetR

$$\frac{\partial r}{\partial t} = Tl \cdot m_r - \left(\left(\frac{\beta\gamma}{Kp + a + r + r_2} \right) + du \right) \cdot r \qquad \text{Eq. 3.2.3-27}$$

$$\frac{\partial m_r}{\partial t} = \frac{Pt \cdot Ts\left(Ka \cdot a \cdot \alpha + 1\right)}{\left(Ka \cdot a + 1\right)\left(Kr \cdot Kd \cdot r^2 + 1\right)^2} - \left(d + du\right) \cdot m_r \qquad \text{Eq. 3.2.3-28}$$

NitR

$$\frac{\partial a}{\partial t} = Tl \cdot m_a - \left(\left(\frac{\delta\gamma}{Kp + a + r + r_2} \right) + du \right) \cdot a \qquad \text{Eq. 3.2.3-29}$$

$$\frac{\partial m_a}{\partial t} = \frac{Pt \cdot Ts\left(Ka \cdot a \cdot \alpha + 1\right)}{\left(Ka \cdot a + 1\right)\left(Kr \cdot Kd \cdot r^2 + 1\right)^2} - \left(d + du\right) \cdot m_a \qquad \text{Eq. 3.2.3-30}$$

We next introduce the dimensionless variables to rescale the concentration of each component and time. Eq. 3.2.3-27, 28, 29, and 30 are substituted with non-dimensional variables as follows,

TetR

$$\frac{\partial y_2}{\partial \tau} = c_4 y_1 - \left(\frac{c_6}{s} + c_1 \right) y_2 \quad \text{Eq. 3.2.3-31}$$

$$\frac{\partial y_1}{\partial \tau} = \frac{\left(\alpha x_2 + 1 \right)}{\left(x_2 + 1 \right)\left(y_2^2 + 1 \right)^2} - c_2 y_1 \quad \text{Eq. 3.2.3-32}$$

NitR

$$\frac{\partial x_2}{\partial \tau} = c_3 x_1 - \left(\frac{c_5}{s} + c_1 \right) x_2 \quad \text{Eq. 3.2.3-33}$$

$$\frac{\partial x_1}{\partial \tau} = \frac{\left(\alpha x_2 + 1 \right)}{\left(x_2 + 1 \right)\left(y_2^2 + 1 \right)^2} - c_2 x_1 \quad \text{Eq. 3.2.3-34}$$

where

$$x_1 = m_a,\ x_2 = Ka \cdot a$$

$$y_1 = m_r,\ y_2 = \sqrt{Kr \cdot Kd}\ \cdot r$$

$$\tau = t \cdot Pt \cdot Ts$$

$$c_1 = \frac{du}{Pt \cdot Ts},\ c_2 = \frac{\left(d + du \right)}{Pt \cdot Ts},\ c_3 = \frac{Ka \cdot Tl}{Pt \cdot Ts},\ c_4 = \frac{\sqrt{Kr \cdot Kd}\ \cdot Tl}{Pt \cdot Ts}$$

$$c_5 = \frac{\delta \gamma}{Pt \cdot Ts},\ c_6 = \frac{\beta \gamma}{Pt \cdot Ts},\ s = Kp + \frac{x_2}{Ka} + \frac{y_2}{\sqrt{Kr \cdot Kd}} + \frac{2 y_2^2}{Kr}$$

These four ODEs from Eq. 3.2.3-31, 32, 33, and 34 are the fundamental framework to describe overall dynamics of the oscillatory system. Underlying new non-dimensional variables rescaling, system behavior depends on $c_1, c_2, c_3, c_4, c_5, c_6$ and α.

In order to visualize this oscillator dynamic behavior, we next plotted x_1-x_2-y_2 3D phase plane (for visualization, we ignored the fourth dimension presented as y_1 because of the dynamics of y_1 identical to x_1) (Figure 11c). The fixed points in 3D phase plane presents the steady state of the system and locate at intersections of the four nullclines when four ODEs (Eq. 3.2.3-31, 32, 33, and 34) equal to 0.

In addition, to further approximate the stabilities of steady states at the fixed points on the phase plane (the intersections of these nullclines), the Jacobian matrix J is

$$J=\begin{pmatrix} -\left(\frac{c_6}{s}+c_1\right) & c_4 & 0 & 0 \\ \frac{-4y_2\left(\alpha x_2+1\right)}{\left(y_2^2+1\right)^3\left(x_2+1\right)} & -c_2 & \frac{\alpha-1}{\left(y_2^2+1\right)^2\left(x_2+1\right)^2} & 0 \\ 0 & 0 & -\left(\frac{c_5}{s}+c_1\right) & c_3 \\ \frac{-4y_2\left(\alpha x_2+1\right)}{\left(y_2^2+1\right)^3\left(x_2+1\right)} & 0 & \frac{\alpha-1}{\left(y_2^2+1\right)^2\left(x_2+1\right)^2} & -c_2 \end{pmatrix} \quad \text{Eq. 3.2.3-35}$$

and the characteristic equation of the system is

$$\det\left(J-\lambda I\right)=\left(\left(\frac{c_6}{s}+c_1\right)+\lambda\right)\left(\left(c_2+\lambda\right)\left(\left(\left(\frac{c_5}{s}+c_1\right)+\lambda\right)\left(c_2+\lambda\right)-\frac{c_3\alpha-c_3}{\left(y_2^2+1\right)^2\left(x_2+1\right)^2}\right)\right)+$$

$$c_4\left(\frac{4y_2\left(\alpha x_2+1\right)}{\left(y_2^2+1\right)^3\left(x_2+1\right)}\left(\left(\left(\frac{c_5}{s}+c_1\right)+\lambda\right)\left(c_2+\lambda\right)-\frac{c_3\alpha-c_3}{\left(y_2^2+1\right)^2\left(x_2+1\right)^2}\right)+\left(\frac{\alpha-1}{\left(y_2^2+1\right)^2\left(x_2+1\right)^2}\right)\left(\frac{4c_3y_2\left(\alpha x_2+1\right)}{\left(y_2^2+1\right)^3\left(x_2+1\right)}\right)\right) \quad \text{Eq. 3.2.3-36}$$

The stability of each of the fixed points from the intersection of the nullclines can be analyzed by solving the eigenvalues λ from characteristic equation. In the dynamical system (differential

equation system), the sign of real part and availabilities of imaginary part of eigenvalues indicates the dynamics of the system in a long period time, such as exponential increase, decay to a fixed point, limit cycle (Strogatz, 2001). The phase diagram of this dynamical system was constructed for [Caprolactam] versus [aTC] plane (Figure 11a). These parameters have capability of being controlled by the experimental design. Therefore, we superimposed the experimental observations to the phase diagram with respect to changes in both inducer concentrations by model predictions (Figure 11a). This method can in general implement in perturbing other parameters in the model and simulating the corresponding system stabilities; the predictions can be further verified by experimental measurements. This new strategy can be used greatly in helping experimental design and evaluation for cell engineering and reprogramming.

3.3 Results

3.3.1 Model construction and data fitting

To understand the interactions of each component in the system and demonstrate the function of the oscillator gene circuit, we first developed a deterministic kinetic model (Ch 3.2.1). This full model described the mechanic fashion of each time-dependent process involved in oscillator circuits shown as Figure 5a, including TF-DNA kinetic binding, transcription, translation, protein dimerization and enzymatic degradation. The model simulation showed oscillatory capability of the designed circuits (Figure 5c). Additionally, the mRNA and protein predicted trajectories of activator and repressor over time showed that time-delayed response of the functional repressor, due to repressor dimerization and alternation of regulator binding affinities by the inducers, is essential to produce self-sustained oscillations. For mode parameterization, all parameters were categorized into three groups. First, only a few parameters are known in mycobacteria (including mRNA degradation rate and plasmid copy number).

Second, a group of kinetic parameters were extracted from the experimental data, in which the reactions functions were decoupled from the model and then fitted under the assumption of steady-state condition (Ch 3.2.2.1 to 3). For the remaining physical parameters, we implemented a Bayesian MCMC algorithm to estimate the posterior distributions of those parameters in the parameters space by fitting the model to the experimentally measured single-cell oscillatory fluorescent signals (Eydgahi *et al*, 2013)(Ch 3.2.2.4). All parameters required in the model were described in Table 2. We implemented the refined model to predict oscillation's periods at varying both concentrations of Cap and aTC and superimposed simulated results to experimental titration data (Figure 10, green circle and filled blue box). The refined model with estimated parameters showed an accurate prediction. Both deterministic and stochastic model agreed tunable behaviors of amplitude and period under inducer concentration titrations. Notice that the alteration of the period under varying aTC (from 220 min to 250 min) was far less than one under Cap titration (from 200 min to 300 min) in both experimental data and model simulations, suggesting that activity of the positive feedback loop in oscillator circuits controlled by Cap concentration served a role to allow tunability of period. This has been demonstrated in previous report (Stricker *et al*, 2008) in which a simplified oscillator with a negative feedback loop only exhibited incapability for tuning period.

3.3.2 Stability analysis of the oscillatory system

Next we studied the stability of this dynamical oscillatory system with respect to changes in both inducer concentrations. In order to reduce model complexity for dynamical system stability analysis, we categorized the kinetic reactions into two classes, fast and slow reactions, in term of time scale of the reaction rate (Ch 3.2.3). Contrast to slow reactions having rate constant with order of magnitude of minutes, the reactions in the fast class with rate of order seconds were

assumed to be in equilibrium (Hasty *et al*, 2001). With this assumption, we were able to simplify the detail model down to a four-dimensional model (4 of ODEs) (Eq. 3.2.3-15 to 18), the dynamical variables of which were mRNA (m_a, m_r) and protein (a, r) concentrations of both activator and repressor. Next in order to reduce the number of parameters involved in the four-dimensional model, we introduced dimensionless variables to linearly rescale the original variables and time. Finally, the four linearized ODEs model was the fundamental framework to describe overall dynamics of the oscillatory system, and its dynamical behavior depends only on dimensionless variables c_1 - c_6, s and α (Eq. 3.2.3-31 to 34).

The stabilities of the fixed points of the dynamical system, the intersection of the nullclines, can be analyzed by solving the eigenvalues of characteristic equation derived from the Jacobian matrix (Eq. 3.2.3-35). To visualize the stabilities which represented long-term dynamical oscillatory behavior, a phase diagram was plotted with respect to changes in the space of both aTC and Cap concentrations (Figure 11a). As results, the phase diagram showed three regimes labeled as: (1) an unstable steady state leads to limit cycle oscillation; (2) decay oscillation to a monostable steady state; (3) exponential decay to a monostable steady state. For each of three classes, we showed the corresponding time-dependent trajectories (Figure 11b) and 3-dimentional x_1 -x_2 -y_2 phase planes (Figure 11c), in which non-dimensional variables x_1, x_2, y_1, y_2 and τ were rescaled from the concentration of activator mRNA, activator protein, repressor protein, and modeling time (min), respectively after nondimensionalization. We selected x_1 -x_2 -y_2 planes because the trajectories of x_1 and y_1 were identical and therefore y_1 variable was ignored. In the first class, only one fixed point (Figure 11c-1; red circles) was identified and example phase portraits exhibited robust limit cycle oscillations from different initial starting points. For the second class, differences of aTC amount determined the final level of monostable steady state

(Figure 11b-2). Regardless, all example phase portraits circulated and finally merged to the fixed point (Figure 11c-2, dark green circles). In the last class, the portraits immediately approached to the steady state without spiraling (Figure 11b-3; Figure 11c-3, dark green circles). To further validate the predictive power of this simplified model, we projected 18 of experimental conclusions from both microfluidics and fluorescence plate reader results on the phase diagram with respect to inducer concentrations (Figure 11a, stars and crosses). Decay oscillation and exponential decay were actually difficult to be distinguished from our low-frequency oscillated data with the period of 24 hours measurement. Therefore, the symbols of star indicated oscillations continuing during measurement, and the symbols of cross then presented the signals approaching to steady state (the signal variations in the last quarter of the measurement points less than the one in all measurement points times the empirical values of 0.1). As results, the results from experiments can be explained by our phase diagram model prediction. The phase diagram indicated that the robustness of oscillator system happened in broad range of inducer concentrations underlying our design principle in mycobacterial condition.

Chapter 4. Probing temporal dynamics of the gene-regulatory network using oscillator perturbations

4.1 Introduction

To couple the synthetic gene oscillator to the mycobacterial regulatory network, the transcriptional regulator *kstR* was translationally fused with a red fluorescence gene and placed under the control of the oscillator circuit (Figure 5a). KstR and its regulon are involved in diverse lipid metabolism, and KstR is required for Mtb infection (Kendall *et al*, 2007). To unveil the temporal dynamics of KstR regulations, the expression profiles of KstR and its directly regulated genes were used as the output of the regulatory system. To do so, two genes (MSMEG1953 and 6038) were selected, which have been experimentally verified as KstR targets by microarray and Chip-Seq (Galagan *et al*, 2013; Kendall *et al*, 2007). The upstream intergenic regions, including promoters, of MSMEG1953 and 6038 genes were transcriptionally fused with a yellow fluorescence gene (Figure 5a). To keep the same gene copy numbers of input and output genes, all the synthetic circuits were placed in an plasmid to diminish the variables may causing gene expression fluctuation (pWOsc6KC1953 and pWOsc6KC6038, Table 1).

4.2 Methods

4.2.1 Build a extended regulatory network coupled with the oscillator circuit

For mycobacteria oscillator extending to regulatory network, the *rfp* gene in the cassette as previously described was replaced by a fused *kst*R-*rfp* gene, in which *kst*R gene without the stop codon were PCR-amplified from *M.Smegmatics* genome. Next several intergenic regions of the targeted genes which were specifically bound by KstR and validated by Chip-Seq results were cloned and fused with yellow fluorescence protein *yfp* gene following by an ssrA degradation tag.

Those KstR regulated gene cassettes were inserted back to the plasmid with the oscillator gene circuits individually (pWOscKC1953 and pWOscKC6038). In order to eliminate KstR regulations from chromosomal gene copy, we constructed *M.Smegmatics* strain with Δ*kst*R in which part of kstR gene locus on chromosome was replaced by a cassette containing Zeocin resistance gene, and enhanced cyan fluorescence protein *ecfp* gene with constitutive promoter. Double cross-over site-specific recombination screening for mycobacterial cells followed the previous report (Muttucumaru and Parish, 2004). Constitutively expressed ECFP during experiments provided a constant background signals for single cell experiment. pWOscKC1953 and pWOscKC6038 were then transformed to *M.Smegmatics* Δ*kst*R.

4.2.2 ODEs of kinetic reactions in extended regulatory network in Mycobacteria

Here we provide one example for proof of principle that connects oscillator to a simple regulatory network in *M. Smegmatics*. Transcription factor KstR was chosen to be driven by oscillator gene circuits, and several target genes in KstR regulon including positive and negative regulations were selected to respond the temporal dynamics of KstR in the cells. In our experimental design, the target gene promoters are fused with fluorescence reporter gene which is tagged by ssrA sequence. So the dynamics of fluorescence reporter gene expression is regulated by KstR as well as protein degradation rate. We assume that KstR binding to target gene promoters and activating/repressing target gene transcription follows cooperative binding which can be described by Hill function (Eq. 4.2.2-1).

$$\mathrm{KstR} \xrightarrow{\alpha \cdot Ts \cdot g(\mathrm{KstR})} \mathrm{KstR} + \mathrm{m}_t$$

$$\text{where } g(KstR) = \left(\frac{1}{1 + \left(\frac{km_k}{KstR} \right)^{h_k}} \right) \quad \text{if} \begin{cases} \text{KstR activates target} & h_k > 0 \\ \text{KstR represses target} & h_k < 0 \end{cases}$$

Eq. 4.2.2-1

$$\mathrm{m}_t \xrightarrow{Tl} \mathrm{m}_t + \mathrm{T}$$

Eq. 4.2.2-2

$$\mathrm{m}_t \xrightarrow{d+du} \varnothing$$

Eq. 4.2.2-3

$$\mathrm{T} \xrightarrow{\kappa \cdot f(\mathrm{ssrA}) + du} \varnothing$$

Eq. 4.2.2-4

Where *g(KstR)* corresponds to Hill function and the sign of hill coefficient h_k indicates either positively or negatively cooperative KstR binding to target. m_t is mRNA of fluorescence reporter gene regulated by KstR and mRNA degradation *d* and dilution *du* rate uf_t and T here represent the reporter gene before and after protein folding and their stability of the protein are also determined by ssrA-driven degradation and dilution rate ($\kappa f(ssrA) + du$). Eq. 4.2.2-2, 3, and 4 presented translation, mRNA degradation, and protein degradation reactions, respectively.

4.3 Results

Using time-lapse fluorescence microscopy of mycobacterial cells growing in microfluidic devices for 24 hrs, dual-color fluorescence time courses (RFP and YFP) were captured at single-cell resolution. In total, we analyzed 376 single-cell input-output trajectories for KstR-MSMEG1953 and 422 trajectories for KstR-MSMEG6038. We also extended the mathematical oscillator model to depict KstR dynamic regulations by implementing a Hill function on the assumption of KstR binding to the promoters following cooperative binding (Ch 4.2.2**,** Eq. 4.2.2-1). After model fitting, the correlated expressions of oscillator-driven KstR and its regulated

gene targets were plotted (Figure 12a). The responses of MSMEG 1953 and 6038 showed distinct dynamic behaviors corresponding to the same KstR input.

We in advance computed their cross-correlation functions between input and output (Dunlop *et al*, 2008)(Figure 12b). The cross correlations extract several characteristics of gene regulation from dynamic input-output profiles: (1) The cross-correlation function is symmetric because of sinusoidal-like input-output signals, and active regulations between transcription factor and target genes appear on maximum/minimum of cross-correlation functions when $d > 0$, and where the time of the point indicates an effective regulation delay time, denoted as d_{reg}. (2) The sign of $C(d_{reg})$ indicates the direction of regulation. KstR-MSMEG1953 generated a delayed correlation, whereas KstR-MSMEG6038 produced a delayed anticorrelation. Remarkably, KstR displays a bifunctional regulation to different target genes, which appears to contradict the previous microarray results that suggested KstR functions as a repressor (Kendall *et al*, 2007). In addition, the difference of delay time d_{reg} between MSMEG1953 and MSMEG6038 implied that KstR binding affinity is dependent on the target promoter sequences. This assumption matched the model parameterization, in that Km_{k1953} (The binding affinity to MSMEG1953) was estimated as 18.71 molecular/fl and Km_{k6038} as 164.36 molecular/fl (Table 2). The cross correlation between KstR and background fluorescence signals was also calculated (Figure 12b, the bottom plot), which was encoded by a constitutively expressed enhance cyan fluorescence gene translationally tagged with a protein degradation tag and integrated into mycobacteria genome. Interestingly, a non-delayed correlation was found, indicating that both *kst*R and *ecfp* were under the same extrinsic (global) regulations such as protein degradation, cell duplication, etc. In this case, protein degradation was the major contribution because KstR and ECFP were conjugated to the same type of ssrA protein degradation tag. Compared to the cross-correlation between KstR and

constitutively expressed ECFP, the temporal responses of MSMEG1953 and MSMEG6038 were exclusively controlled by the dynamics of KstR expression.

Chapter 5. Discussion and conclusions

5.1 Discussion

We have shown to use synthetic biology combining with computational modeling to engineer mycobacterial cells and in advance probe the temporal dynamics of gene regulations underlying a complicated gene regulatory network. The theoretical design of the oscillator which comprise a positive and a time-delayed negative feedback loop (Hasty *et al*, 2002; Smolen *et al*, 1998) successfully generated autonomous self-stained oscillatory gene expression in mycobacterial cells (this report), *E.coli* (Stricker et al, 2008) and mammalian cells (Tigges *et al*, 2009), suggesting that this design framework may also generally apply into other organisms. In addition to the framework design, protein degradation rate of the gene circuits also contributes to reshaping dynamic profiles of gene expression (Eldar *et al*, 2010; Hasty *et al*, 2002; Hasty *et al*, 2001). In the case that if the protein lifetime is longer than oscillatory period after proteins being translated, the accumulation of proteins over time tends to average out the pattern of fluctuation generated by the gene circuits. Therefore, adding an appropriate degradation system is able to reduce time-averaging effect. In mycobacteria system, SsrA RNA encoding a 11-amino acids (AANDENYALAA) added on the C-terminal end of the protein for protease targeting had been reported (Karzai *et al*, 2000). To evaluate ssrA tag degradation efficiency, we first translationally tagged this 11 amino acid peptide on the C-terminal sequences of NitR, and RFP in presence of the positive feedback loop only (pNG1, Table 1). The fluorescence signals kept at equilibrium after Cap adding (Figure 2, left plot), suggesting that the protein degradation rate seemed to be equal to protein synthesis rate. In order to still capture temporal dynamics, we engineered the amino acid sequence by individually substituting one of the residues (L9A and A11D) to slow down degradation rate by reducing the specificity of protease recognition (Chang *et al*, 2008;

Flynn *et al*, 2001). The two modified ssrA tags with medium and slow degradation respectively relative to unmodified tag were tested in the same gene circuit (pNG2; pNG3, Table 1). Consequently, the positive feedback loop exhibited a sensitive regulation corresponding to Cap concentration in 4 hours measurement for both medium and slow degradation tags (Figure 2, middle and right plot).

The design principle of the chimera promoter ($P_{nit/tet}$) in oscillator gene circuit theoretically exhibited a logic combinational regulation in which genes are actively expressed only when NitR ∩ none-TetR on the chimera promoter. This type of the behavior critically helps generate a time-delayed negative feedback loop in presence of both Cap and aTC (Figure 5b). In order to experimentally validate this theoretical design, we analyzed the reporter fluorescence expression of the oscillator gene circuits (pWOsc4, Table 1) under three conditions with different inducer concentrations (Figure 3). In the first condition in absence of Cap and aTC, a few of NitR and TetR proteins were only available in the cells; however NitR kept in inactive form whereas TetR was active. In 25 hours of time-series measurements, the fluorescent signals represented as oscillator gene expression showed in a basal-line-like flat line (Figure 3, green line), indicating that the circuits were continuously shot down. In the second condition that only Cap available, both NitR and TetR were considered as active form because of NitR conformationally changing to active form when Cap bound. The fluorescence signals showed slightly higher than the precious condition (Figure 3, red line), suggesting that TetR bound to the chimera promoter still effectively blocked RNA transcription, even when active NitR existed. In the last condition when both inducers available, TetR converted to inactive form and NitR kept the same active form. The fluorescence then dramatically increased relative to a baseline (Figure 3, blue line), suggesting that inactive TetR no longer blocked the accessibility of RNA polymerase and active NitR

activated the fluorescence reporter gene expression. Consequently, the experimental data showed a cohesive result as the theoretical design predicted.

Using quantitative information about individual cells extracted from time-series images, we also analyzed temporal patterns of cell morphology and traced cell lineages at single-cell resolution. For each ancestor namely that the first cells appeared at time zero images, we reconstructed cell lineages and connected to cell-to-cell heterogeneity in term of cell size, growth rate and corresponding fluorescence intensity trajectory driving by the oscillator gene circuits between daughter cells, which were representatively shown in Figure 14. In summarized results, single cells display remarkable heterogeneity in growth rate and cell doubling time. Combined with the phenotypic features of mycobacterial cell in asymmetric division (Aldridge *et al*, 2012; Joyce *et al*, 2012; Santi *et al*, 2013), we tried to correlate cell size at birth to growth rate and doubling time (Figure 15a, b). Strikingly, cell size at birth positively correlated to growth rate (absolute cell size increase over time at birth to next division), but negatively correlated to doubling time. Additionally, fast-growing cells have shorter time for duplication than slow-growing cells (Figure 15c). Recently two reports in the study of cell-growth controlling mechanism in mycobacteria suggested correlated but contradictory theories of cell division between time-dependent (Aldridge *et al*, 2012) and size-dependent manner (Santi *et al*, 2013). Our observations tend to support a second scenario in which the cell division event in mycobacteria may be mainly regulated by cell size. However, we also have similar observations as the report (Santi *et al*, 2013) that large cell size at birth attempted to reach larger cell size at division than small cell size at birth, suggesting that not only cell size (Figure 15d), but some other mechanisms may exist to control cell dividing event. Regardless, these dynamic patterns of

cell morphology at single-cell level provide great insight into a much-debated problem about cell-size homoeostasis in cell biology.

Both deterministic and stochastic model underlying the theoretical design of oscillator gene circuits displayed a robust oscillation in a broad range of inducer concentrations. However, time-lapse microscopic images actually showed that 70 % of individual cells on average in each experiment, which included oscillator alone and oscillator coupled with the network, showed robust oscillation in 24 hours measurement. Other cells might express decayed, steady-state fluorescent signals, and even no fluorescence observed. There are several characteristics in mycobacteria which didn't consider in the mathematical model and might account for generating heterogeneous cellular responses in oscillated gene expression in isogenic cell population. As we discussed previously, cell-to-cell heterogeneities in cell size at birth, growth rate, doubling time have been observed in mycobacterial cells. Furthermore, when analyzed single-cell fluorescence trajectories with the framework of cell lineage, we noticed that some of two descendants exhibited distinguished fluorescence trajectories right after the mother cell dividing (Figure 14, bottom plot), suggesting that protein abundance presented by fluorescence intensity might not be equally distributed during the spatial organization of cellular components during cell cycle. Those observations implied fluctuations and variability in cell-specific factors such as the activity of transcription and translation, metabolite concentrations, cell age, cell size and stage of cell cycle in the population. Consequently, such factors contributed to extrinsic noise which arose from intercellular differences (Elowitz *et al*, 2002), and leaded to unsynchronized oscillation and variance in oscillatory period and amplitude in cell population (Kaern *et al*, 2005). Another consideration for inconsistent oscillation across the cell population may be due to gene copy number of oscillator gene circuits. In our experimental design, the circuits which were comprised

three modules and target genes, if it coupled with oscillator, were carried in a plasmid with 3 to 5 copies in the mycobacterial cells (Stover *et al*, 1991). With comparison of the plasmids used in *Ecoli* oscillator in which had more than 50 copies and expressed a robust oscillation across the population (> 99% of cell population) (Stricker *et al*, 2008), the gene copy number may play a causative factor in variation of oscillation among individual cells. In the previous reports, gene expression noise namely that the distribution of protein levels in individual cells was owed to the low copy number (1-2 per cell) of most genes and dominated by bursts of mRNA productions (Eldar *et al*, 2010; Swain *et al*, 2002). Eventually gene expression noise leads to cell-to-cell variation. To decrease the noise, in addition to increase protein lifetime (Swain *et al*, 2002), reduce translation rate and raise transcription rate (Maamar *et al*, 2007), increasing the number of gene copies has also been used in the case of *Bacillus subtilis* cells (Suel *et al*, 2007). Furthermore, it is still not clear about the mechanism of plasmid partition in mycobacteria, and the robustness of duplicating plasmids distributed into descendants during cell dividing. Variance in random distribution of the plasmid with low copy number during partitioning should be higher than the one with high copy number as expected intuitively. With all these information summarized, low copy number of plasmid might be the potential element for generating cell-to-cell heterogeneous gene expressions driving by oscillator gene circuits.

Here we not only developed a new engineered approach to exploit the temporal dynamics of the gene regulatory network in mycobacteria, but provided a promising way that utilize computational modeling with dynamical systems theory to guide the experimental designs for building the gene circuits with a specific and programmable functional behavior, particularly in mycobacteria. In the field of drug therapy to latent TB infection, there are currently many active researches on accessibility of the drug to the target and shortening the courses for treatment

(Denholm and McBryde, 2010). This synthetic biology-based approach combined with systems biology study provides a new framework to re-engineering Mtb and Bacillus Calmette–Guérin (BCG), which will lead to development of new therapies.

5.2 Conclusion

This report displayed an engineering-based approach to understand temporal dynamics of gene regulatory network in *Mycobacterium Tuberculosis* by using a synthetic and tunable oscillator functioning as a circadian clock for transcriptional network perturbation. Temporal dynamics of gene expression and cell morphology at single-cell resolution were quantitatively captured by implementing microfluidic chip with time-lapse microscopy. This proof-of-concept sets the stage for advance applications in reverse-engineering the regulatory network by measuring global cellular responses of the genome to an oscillatory perturbation. Additionally, the framework of this synthetic biology-based approach combined with systems biology study opens many opportunities to re-engineer Mycobacteria for further drug development.

Tables

Table 1. The strain used in this report

Strain/plasmid	Relevant characteristics	Reference or source
Strains		
M.Smegmatics mc^2 155	WT	(Snapper *et al*, 1990)
M.Smegmatics KZ	Δ(kstR::Zeocin^r)	This study
M.Smegmatics KZEC	Δ(kstR::ECFP-Zeocin^r)	This study
E.coli Top10	General cloning strain	Invitrogen
Plasmids		
pMV306Hyg	Hyg^r	(Stover *et al*, 1991)
pMV261Kan	Kan^r	(Stover *et al*, 1991)
pNG1	pMV261Km derivative, $P_{nit/tet}$-NitR-ftag[1], $P_{nit/tet}$-GFP-ftag	This study
pNG2	pMV261Km derivative, $P_{nit/tet}$-NitR-mtag[1], $P_{nit/tet}$-GFP-mtag	This study
pNG3	pMV261Km derivative, $P_{nit/tet}$-NitR-stag[1], $P_{nit/tet}$-GFP-stag	This study
pT1	pMV306Hyg derivative, $P_{nit/tet}$-TetR-ftag	This study
pT2	pMV306Hyg derivative, $P_{nit/tet}$-TetR-mtag	This study
pT3	pMV306Hyg derivative, $P_{nit/tet}$-TetR-stag	This study
pWOsc1	pMV261Km derivative, $P_{nit/tet}$-NitR-ftag, $P_{nit/tet}$-TetR-ftag, $P_{nit/tet}$-RFP-ftag	This study
pWOsc2	pMV261Km derivative, $P_{nit/tet}$-NitR-ftag, $P_{nit/tet}$-TetR-mtag, $P_{nit/tet}$-RFP-ftag	This study
pWOsc3	pMV261Km derivative, $P_{nit/tet}$-NitR-ftag, $P_{nit/tet}$-TetR-stag, $P_{nit/tet}$-RFP-ftag	This study
pWOsc4	pMV261Km derivative, $P_{nit/tet}$-NitR-mtag, $P_{nit/tet}$-TetR-ftag, $P_{nit/tet}$-RFP-mtag	This study
pWOsc5	pMV261Km derivative, $P_{nit/tet}$-NitR-mtag, $P_{nit/tet}$-TetR-mtag, $P_{nit/tet}$-RFP-mtag	This study
pWOsc6	pMV261Km derivative, $P_{nit/tet}$-NitR-mtag, $P_{nit/tet}$-TetR-stag, $P_{nit/tet}$-RFP-mtag	This study
pWOsc7	pMV261Km derivative, $P_{nit/tet}$-NitR-stag, $P_{nit/tet}$-TetR-ftag, $P_{nit/tet}$-RFP-stag	This study
pWOsc8	pMV261Km derivative, $P_{nit/tet}$-NitR-stag, $P_{nit/tet}$-TetR-mtag, $P_{nit/tet}$-RFP-stag	This study
pWOsc9	pMV261Km derivative, $P_{nit/tet}$-NitR-stag, $P_{nit/tet}$-TetR-stag, $P_{nit/tet}$-RFP-stag	This study
pWOscKstR	pMV261Km derivative, $P_{nit/tet}$-NitR-mtag, $P_{nit/tet}$-TetR-stag, $P_{nit/tet}$-KstR-RFP-mtag	This study

pWOscKC1953	pMV261Km derivative, $P_{nit/tet}$-NitR-mtag, $P_{nit/tet}$-TetR-stag, $P_{nit/tet}$-KstR-RFP-mtag, P_{M1953}[2]-YFP-stag	This study
pWOscKC6038	pMV261Km derivative, $P_{nit/tet}$-NitR-mtag, $P_{nit/tet}$-TetR-stag, $P_{nit/tet}$-KstR-RFP-mtag, P_{M6038}[2]-YFP-stag	This study

[1]. ftag: fast degradation tag; mtag: medium degradation tag; stag: slow degradation tag.
[2]. P_{M1953}: Intergenic region upstream of MSMEG 1953; P_{M6038}: Intergenic region upstream of MSMEG 6038.
[3]. P_{mop}: constitutive promoter (Ehrt *et al*, 2005).

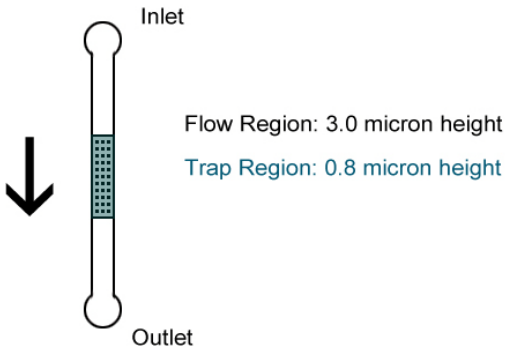
Table 2. The parameters used in this oscillator-extended regulatory network model

Parameters	Description	Values	Comment/Reference
Ts	Transcription	0.36 (min^{-1})	Estimated by MCMC
α	Fold-change of active transcription	20	Estimated by MCMC
Tl	Translation	5 (min^{-1})	Estimated by MCMC
k_d	Protein dimerization	0.0018 (min^{-1})	Estimated by MCMC
k_{-d}	Protein de-dimerization	0.000018 (min^{-1})	Estimated by MCMC
d	mRNA degradation	0.18 (min^{-1})	*Msmeg* data (Rustad et al 2012)
Υ_f	Fast protein degradation	235.43 (molecular/min)	Calculated from fluorescence induction data
Υ_m	Medium protein degradation	71.78 (molecular/min)	Calculated from fluorescence induction data
Υ_s	Slow protein degradation	34.21 (molecular/min)	Calculated from fluorescence induction data
du	Cell dilution	0.002 (min^{-1})	Calculated from single-cell microscopy data
k_{-a}	NitR dissociation rate constant	1.8 (min^{-1})	Estimated by MCMC
B^a_{max}	Maximum of NitR binding affinity	1.02 ($molecular^{-1}$)	Calculated from Pandey et al 2009 induction data
B^a_{min}	Minimum of NitR binding affinity	0.0014 ($molecular^{-1}$)	Calculated from Pandey et al 2009 induction data
Km_a	NitR conc. to reach half of binding affinity	18.8 (mM)	Calculated from Pandey et al 2009 induction data
h_a	Hill coefficient for NitR	3.3	Calculated from Pandey et al 2009 induction data
k_{-r}	TetR dissociation rate constant	1.8 (min^{-1})	Estimated by MCMC
B^r_{max}	Maximum of TetR binding affinity	0.98 ($molecular^{-1}$)	Calculated from Ehrt et al 2005 induction data
B^r_{min}	Minimum of TetR binding affinity	0.08 ($molecular^{-1}$)	Calculated from Ehrt et al 2005 induction data
Km_r	TetR conc. to reach half of binding affinity	22.7 (ng/ml)	Calculated from Ehrt et al 2005 induction data
h_r	Hill coefficient for TetR	3	Calculated from Ehrt et al 2005 induction data
Pt	Gene copy number	3	*Msmeg* data
Km_{k6038}	KstR conc. to reach half of binding affinity to MSMEG6038 promoter	164.36 (molecule/fl)	Estimated by MCMC
h_{k6038}	Hill coefficient for KstR to MSMEG6038 promoter	-3.94	Estimated by MCMC

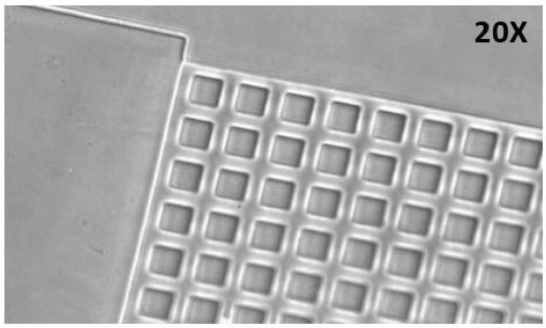
Km_{k1953}	KstR conc. to reach half of binding affinity to MSMEG1953 promoter	18.71 (molecule/fl)	Estimated by MCMC
h_{k1953}	Hill coefficient for KstR to MSMEG1953 promoter	3.94	Estimated by MCMC

Figures

(a)



(b)



(c)

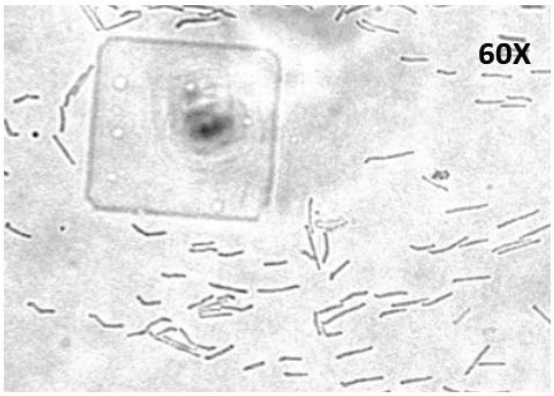


Figure 1. Schematic diagram of microfluidic devices and microscopic images. The microscopic images were from 20X and 60X- magnification of the brightfield objective. (a) Top-down schematic diagram of the microfluidic device showing inlet and outlet holes, edia flow, channel regions, and cell trap regions. Arrow denotes direction of media flow. (b) Brightfield microscope image of the trap region at 20x magnification without cells. (c) Brightfield microscope image of the trap region at 60x magnification without cells.

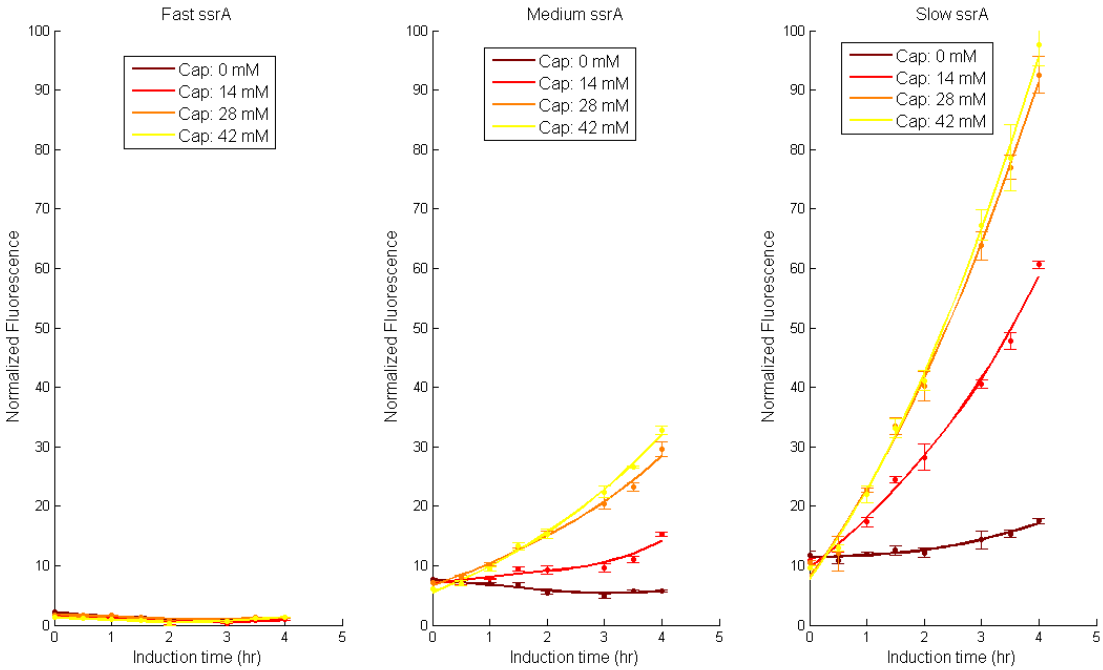


Figure 2. The response of NitR regulation to inducer concentration and degradation tags. The combinatoral tests of ssrA tags on NitR and GFP (the reporter gene) and caprolactam concentration titration on the construct (pNitR-GFP).

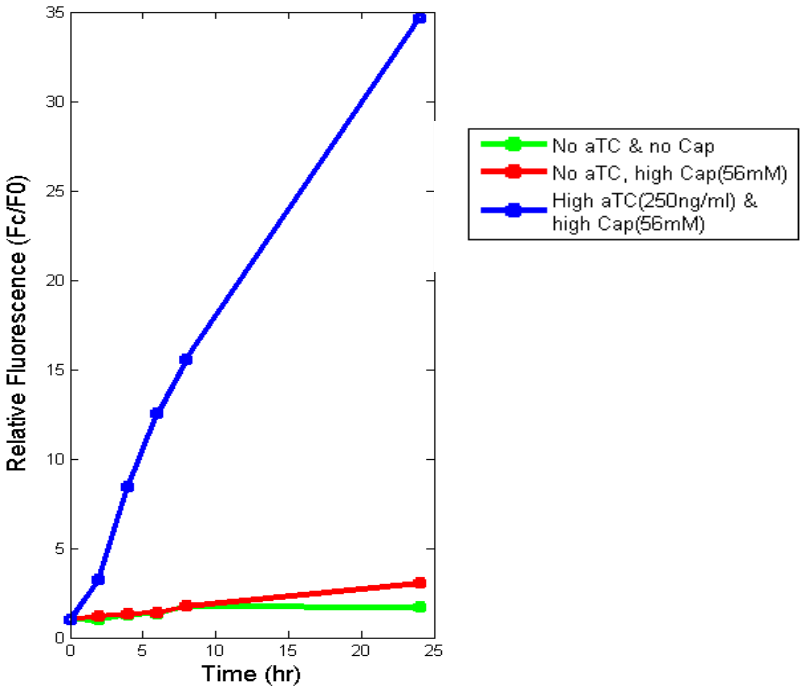


Figure 3. The analysis of NitR and TetR combinational regulation. The fluorescence protein expression under control of hybrid promoter $P_{nit/tet}$ showed the logic regulations of NitR and TetR mediated by Caprolactam and aTC individually. The fluorescent signals were measured from cell population culture by plate reader.

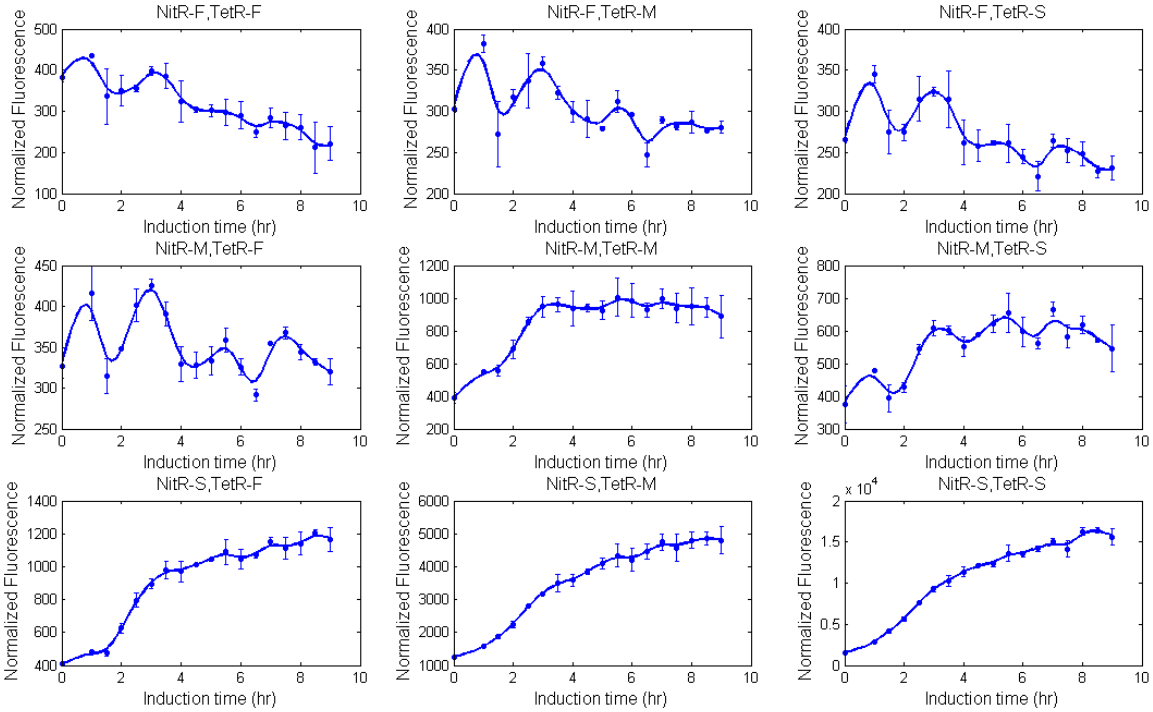


Figure 4. The systematic analysis of oscillation by tagging different protein degradation rates on NitR and TetR. The fluorescence driven by oscillator gene circuit were measured by plate reader every 30 min as sampling frequency. Error bar indicated the standatd deviation of repeated fluorescent signals.

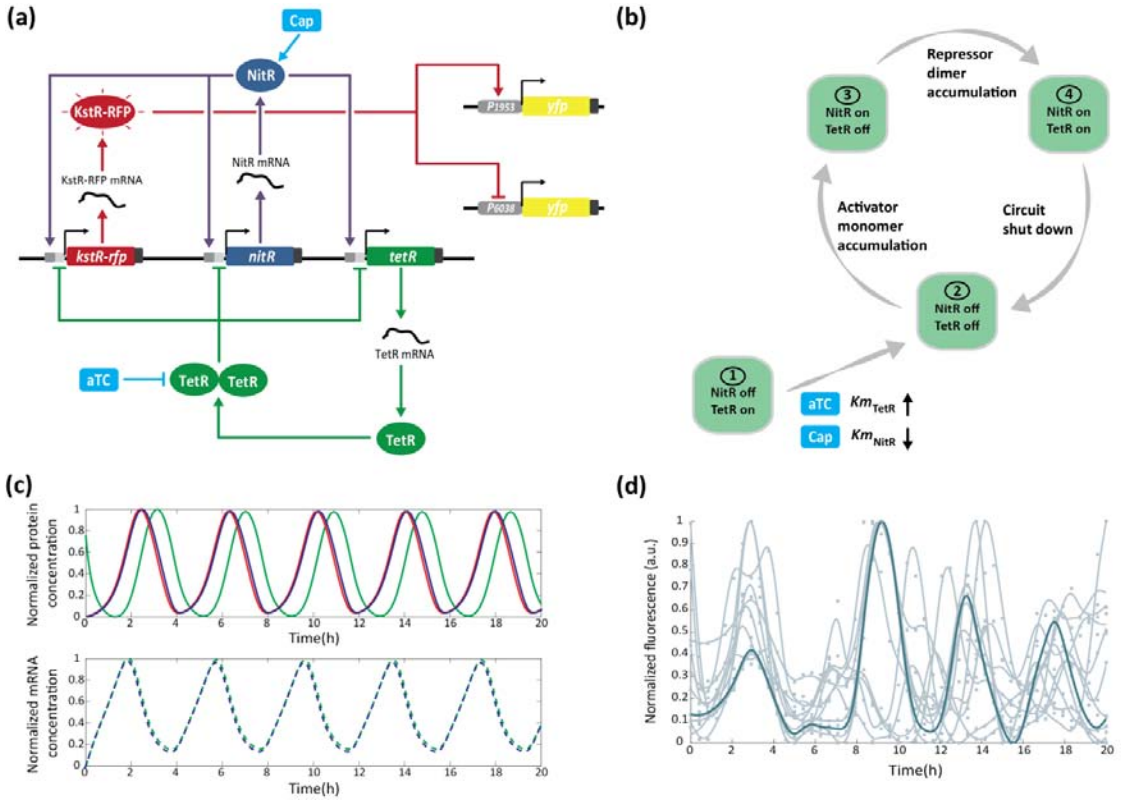


Figure 5. Oscillation in Mycobacteria theoretically and experimentally. (a) The diagram of genetic oscillator gene circuits. The oscillator circuits were composed of three parts: positive (NitR), negative (TetR) feedback loop and the reporter *rfp* gene, and were individually driven by a *nit*A/*Tet*O hybrid promoter. The oscillator was further coupled with mycobacterial regulatory network by connecting to one of TFs-KstR which regulates *yfp* gene under KstR-binding promoter. (b) The mechanism for generating oscillation. The green box indicated the occupancy of NitR and TetR on the hybrid promoters. Adding Cap and aTC inducers leaded to changes in both TetR and NitR binding affinities. Active NitR and TetR differentially binding to hybrid promoter triggered the formation of oscillatory gene expression. (c) The model simulation for defined parameter sets with mRNA concentration (dashed green, TetR; dashed blue, NitR) and protein concentration (red, RFP; blue, NitR; green, TetR dimer). (d) The single-cell fluorescent

trajectories with 25 ng/ml aTC and 14 mM Cap. Gray lines indicated single-cell oscillatory behaviors over time after measured fluorescence intensities (gray points) were normalized by cell size and noise-filtered by a smoothing spline.

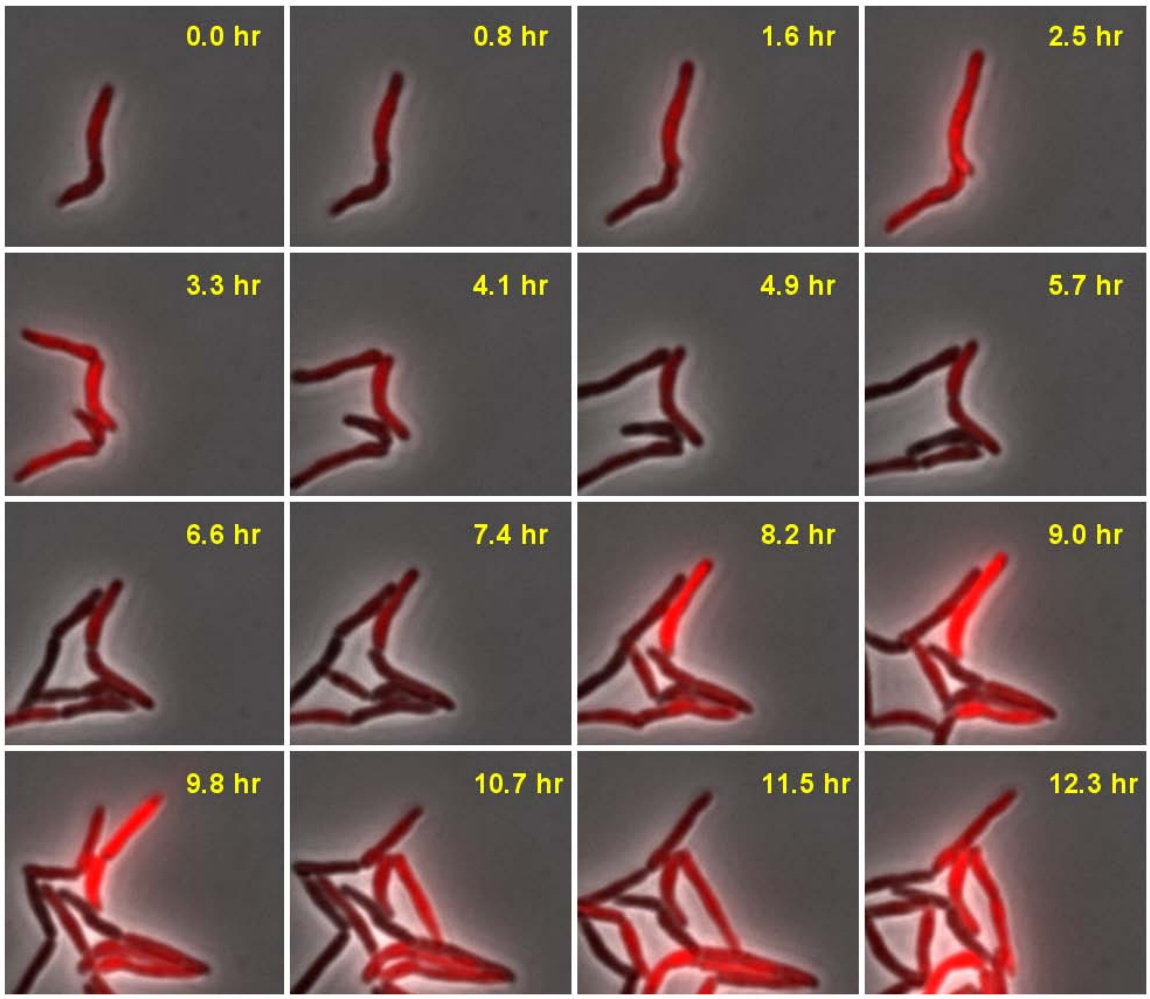


Figure 6. Representative image series of cell oscillatory gene expressions. The red fluorescence intensities representative of oscillator gene circuits by *rfp* reporter gene were superimposed on the phase-contrast images. Numbers indicated measurement hours.

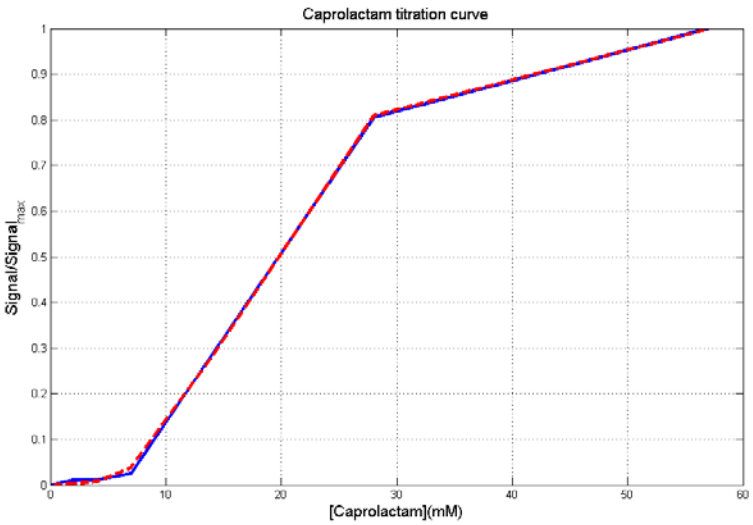


Figure 7. Cap titration data for model parameterization. The caprolactam titration curve from experimental data (Pandey 2009) (blue solid line) and simulations after data fitting to Hill function (red dash line).

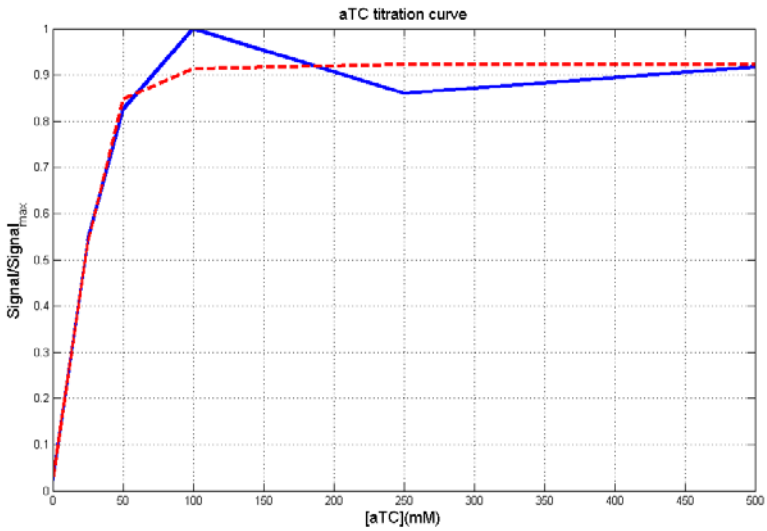
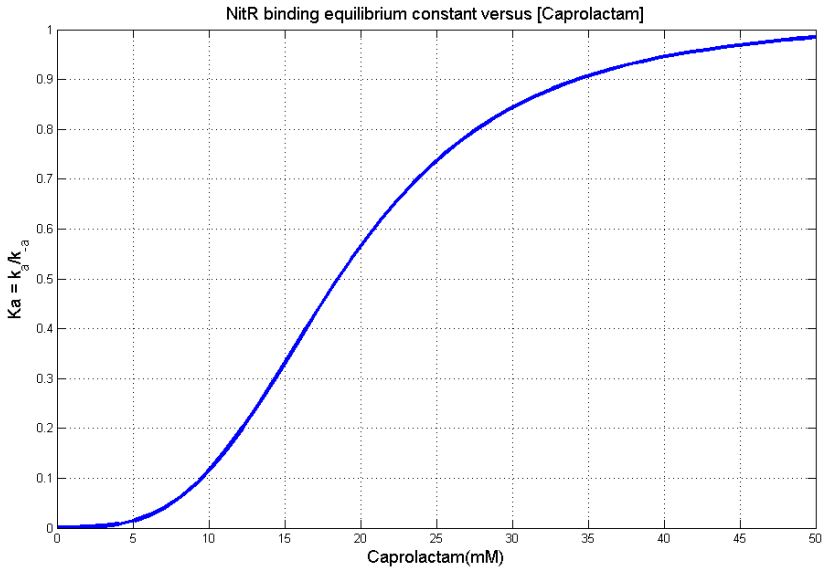


Figure 8. aTC titration data for model parameterization. The aTC titration curve from experimental data (Ehrt 2005) (blue solid line) and simulations after data fitting to Hill function (red dash line).

(a)



(b)

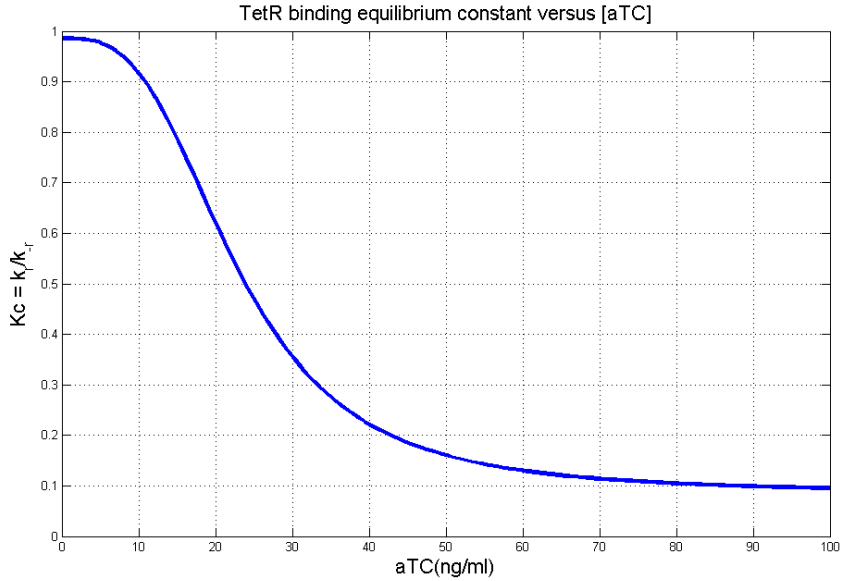


Figure 9. The cooperative regulation of NitR and TetR presented as the model-predicteded Hill functions. (a) NitR *Kc* from the model as a function of [Caprolactam] and [aTC]. (b) The correlation of TetR *Kc* and aTC concentration.

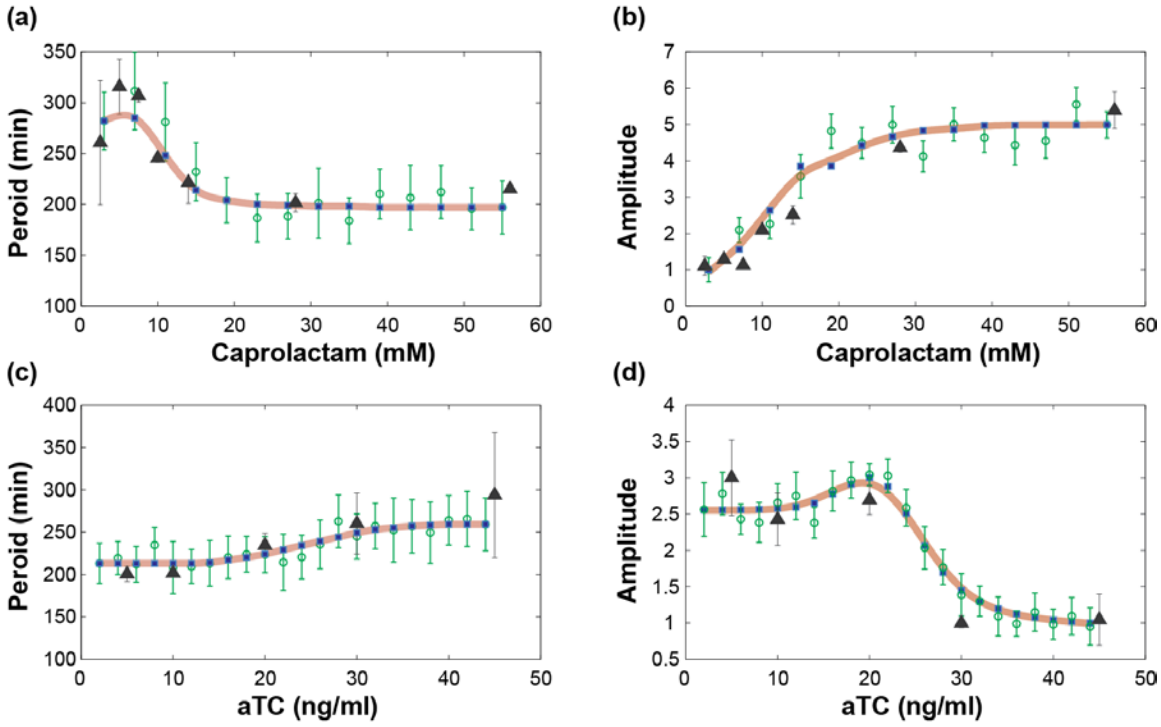


Figure 10. Tunable oscillatory gene expression in mycobacteria. Oscillatory period (a) and amplitude (b) were analyzed with a fixed value of 20 ng/ml aTC and varying Cap. (c) and (d) were under the same condition but fixed Cap concentration at 14 mM and varied aTC. Black triangles with error bars (mean ± s.d.) indicated experimental results measured by fluorescence microplate reader. Blue boxes were the predictions by deterministic modeling and green circles presented mean of repeated stochastic simulations with error bars as s.d. Pink curves showed trend lines based on computational modeling.

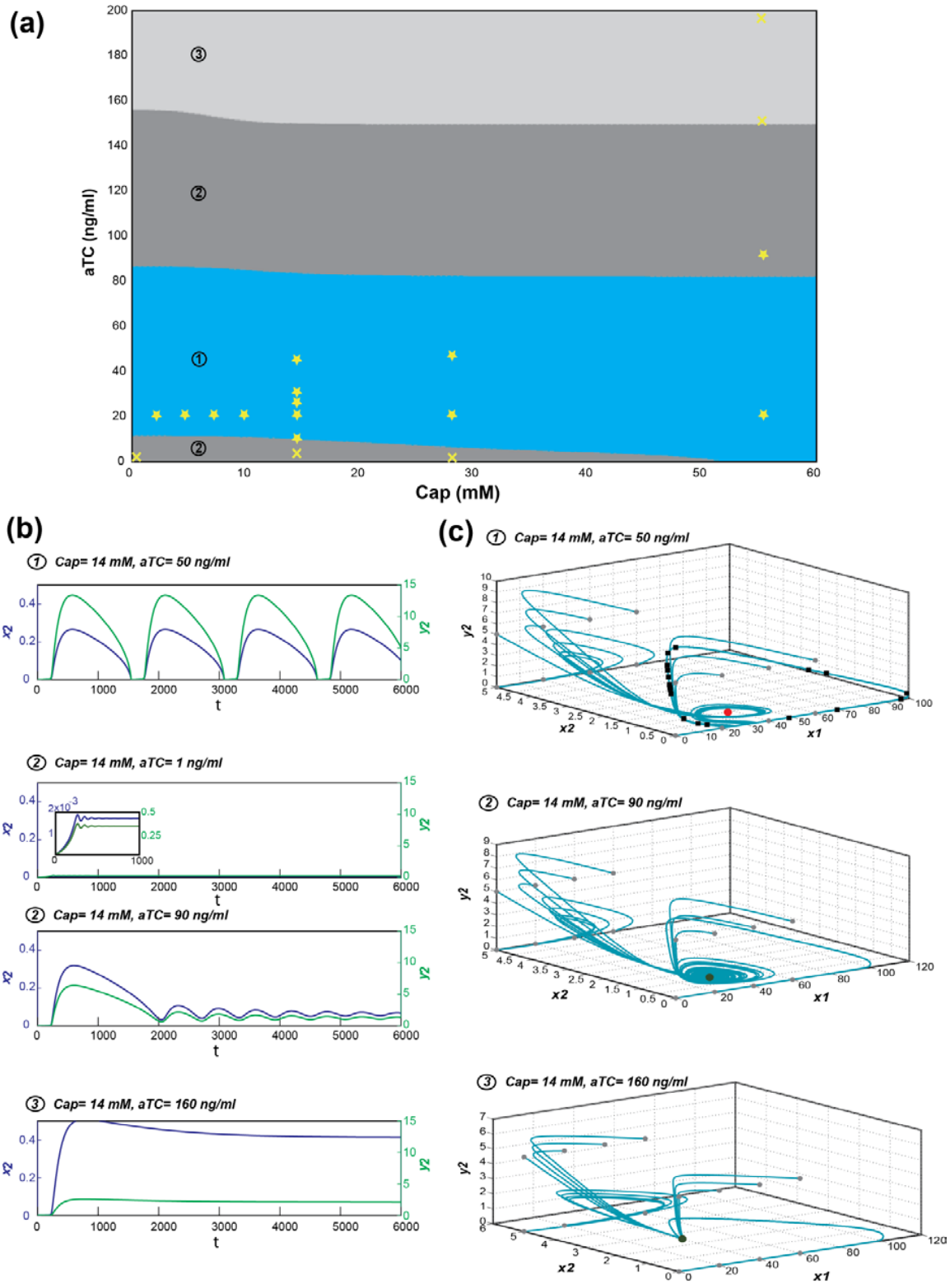

(a)
aTC (ng/ml)
Cap (mM)
(b)
① Cap= 14 mM, aTC= 50 ng/ml
② Cap= 14 mM, aTC= 1 ng/ml
② Cap= 14 mM, aTC= 90 ng/ml
③ Cap= 14 mM, aTC= 160 ng/ml
x2
y2
t
(c)
① Cap= 14 mM, aTC= 50 ng/ml
② Cap= 14 mM, aTC= 90 ng/ml
③ Cap= 14 mM, aTC= 160 ng/ml
x1
x2
y2

Figure 11. Phase diagram and three dimensional phase planes of dynamical oscillator system stability. A phase diagram (a) showed system stabilities with respect to changes in both aTC and Cap concentrations. Three regimes of dynamical behaviors highlighted by different colors were identified and each of which was presented by time-dependent trajectories (b) and three dimensional phase planes (c). x_1, x_2 , y_2 and τ were unitless variables proportional to [activator mRNA], [activator protein], [repressor protein], and modeling time (min), respectively. For time-dependent trajectories, x_2 and y_2 versus τ were plotted by given four different inducer conditions. Each of which were represented by x_1 -x_2 -y_2 three dimensional phase planes. The initial and last states of the trajectories were also shown as gray circles and black boxes. The fixed points where the nullclines interest were marked as red and bark green circles which correspondingly indicated unstable and stable points. In a given range of inducer titrations, the system exhibited (1) limit cycle oscillation; (2) damped oscillation; (3) exponential decay to a monostable steady state. Experimental results were mapped to the phase diagram and shown as ‘oscillating’ (stars) matching to (1) and ‘stable’ (crosses) including (2) and (3).

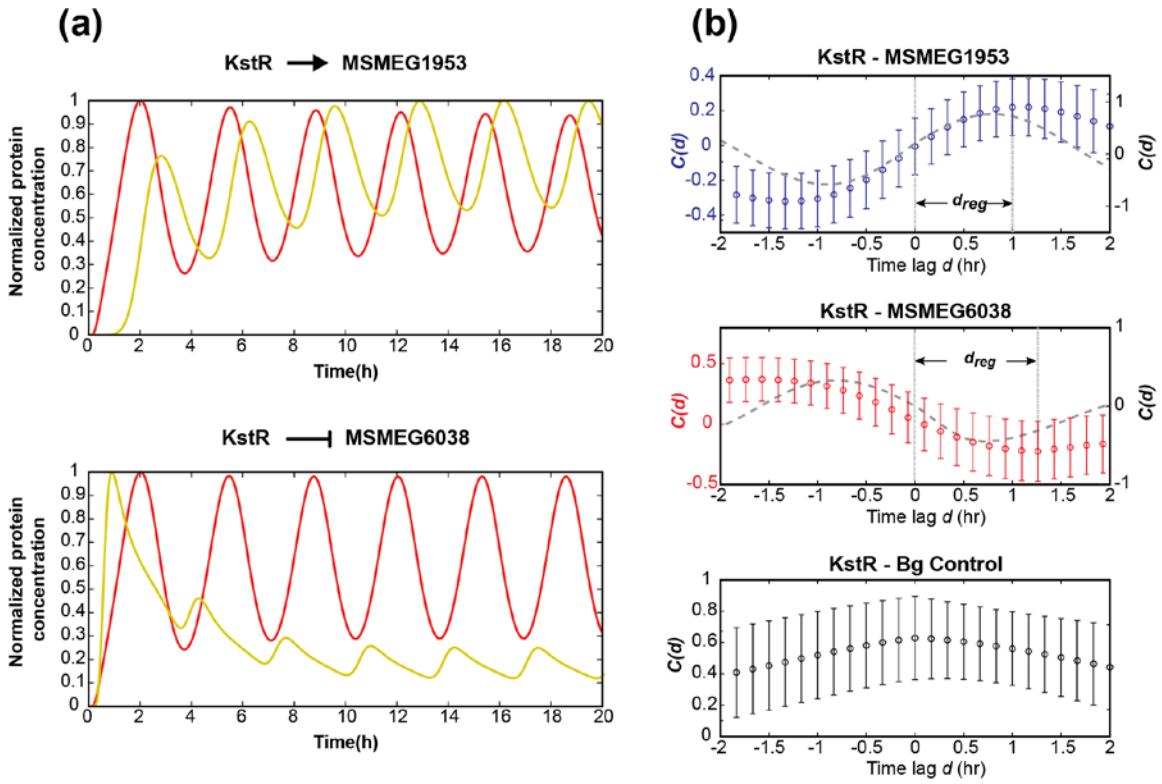


Figure 12. Temporal dynamic regulations of KstR and the target genes. (a) Computational simulations of the temporal dynamics of KstR (red lines) and the target genes (yellow lines) after model parameterization from single-cell input-output fluorescence data. (b) Cross correlation between the temporal profiles of KstR and the target genes. Blue and red circles with error bars (mean ± s.d.) indicated correlation coefficients of KstR -MSMEG1953 and KstR –MSMEG6038 over entire single-cell input-output trajectories. Black circles with error bars showed the correlation between KstR and a constitutively expressed ECFP gene. The gray dash lines presented cross-correlations based on the trajectory simulations in (a).

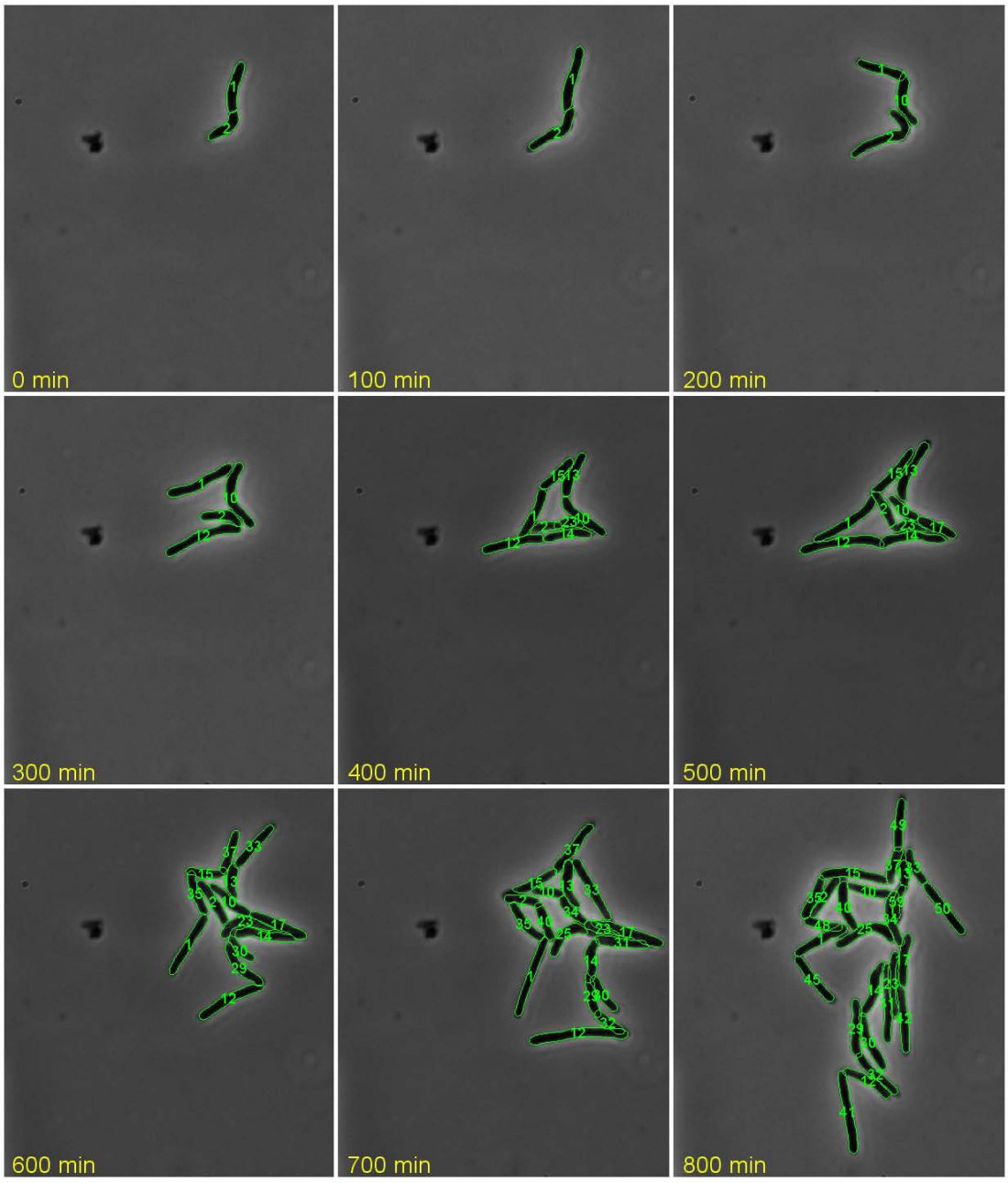


Figure 13. Representative image series of cell boundaries and tracing numbers superimposed on phase-contrast images. The numbers on the individual cells were used to trace cell lineage over series-time images after automatic image processing pipeline.

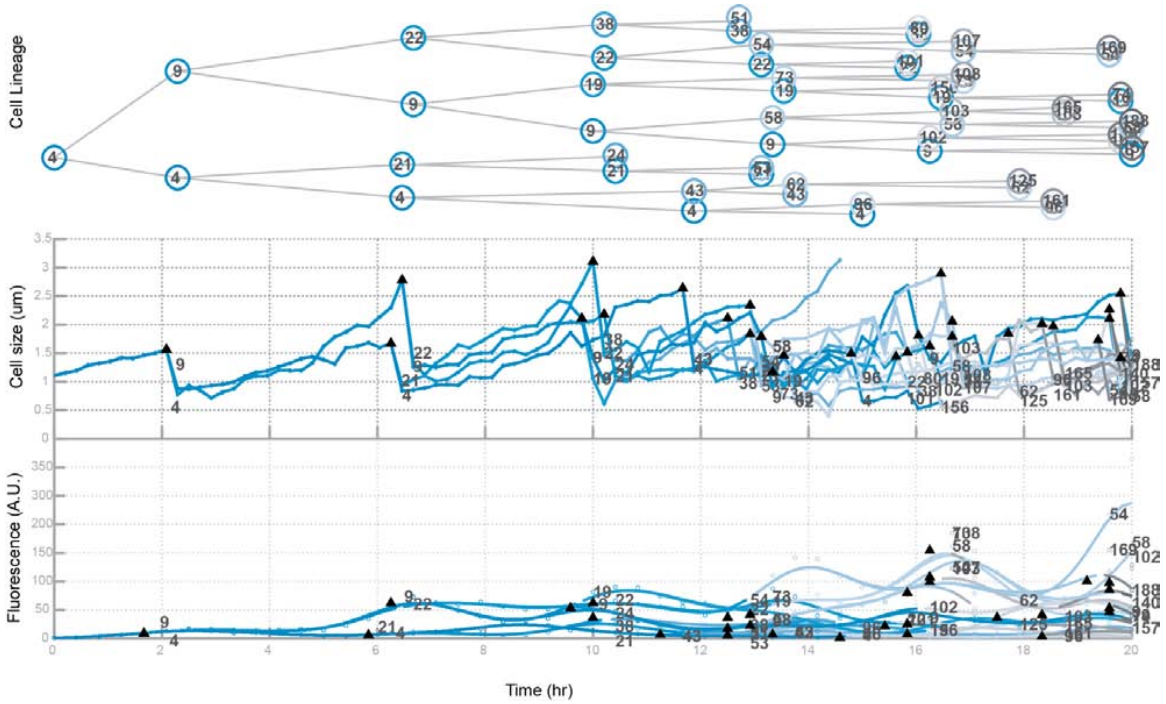


Figure 14. Phenotypic characteristics of individual cells extracted by the software. The dynamic profiles of single-cell gene expression, cell size, position and lineage were extracted according to the coordination of cell boundaries identified on each phase-contrast images and superimposed to the fluorescence images.

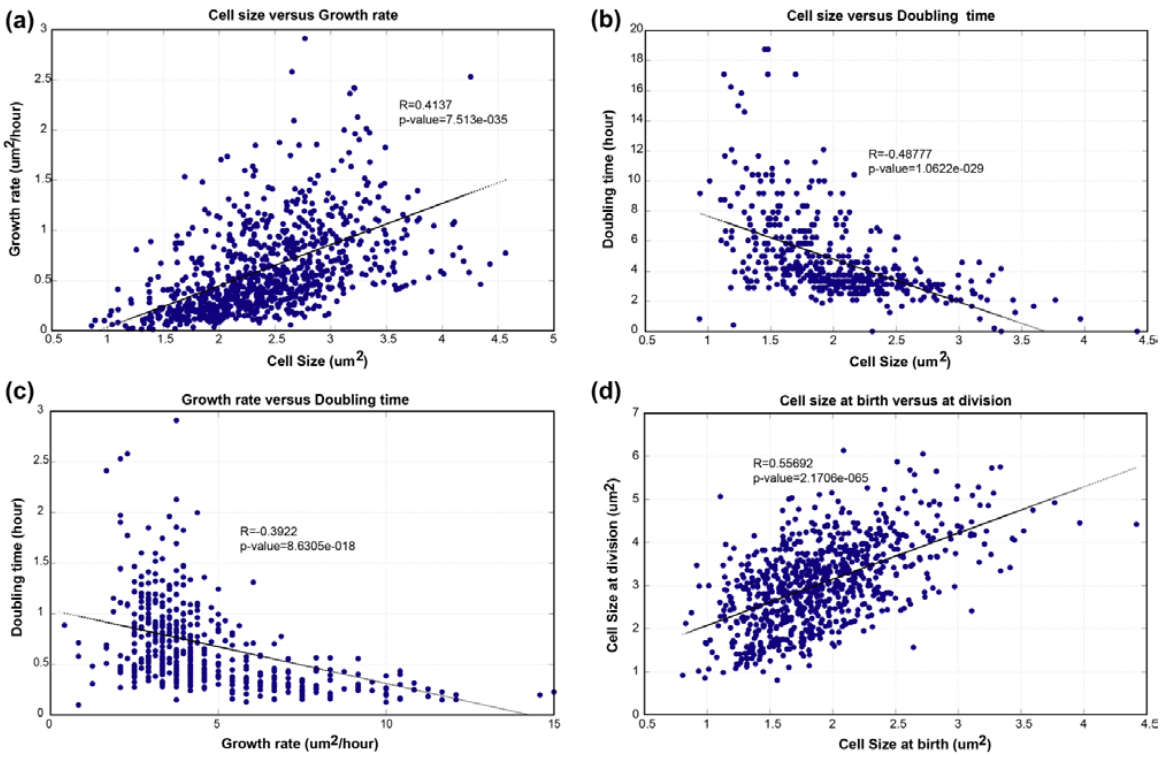


Figure 15. The relationship of cell size at birth and division, growth rate, and doubling time in mycobacterial cells. Cell size at birth was compared to growth rate (a) and doubling time (b). Doubling time was also the function of growth rate (c). Finally, we also compared between cell size at birth and at division (d).

Bibliography

Aldridge BB, Fernandez-Suarez M, Heller D, Ambravaneswaran V, Irimia D, Toner M, Fortune SM (2012) Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science* **335:** 100-104.

Alon U (2007) Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8:** 450-461.

Bennett MR, Hasty J (2009) Microfluidic devices for measuring gene network dynamics in single cells. *Nature Reviews Genetics* **10:** 628-638.

Borrego S, Niubo E, Ancheta O, Espinosa ME (2000) Study of the microbial aggregation in Mycobacterium using image analysis and electron microscopy. *Tissue Cell* **32:** 494-500.

Cai L, Dalal CK, Elowitz MB (2008) Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455:** 485-490.

Chang Y, Wesenberg GE, Bingman CA, Fox BG (2008) In vivo inactivation of the mycobacterial integral membrane stearoyl coenzyme A desaturase DesA3 by a C-terminus-specific degradation process. *Journal of Bacteriology* **190:** 6686-6696.

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics*.

Denholm JT, McBryde ES (2010) The use of anti-tuberculosis therapy for latent TB infection. *Journal of Infection and Drug Resistance* **3:** 63-72.

Dunlop MJ, Cox RS, 3rd, Levine JH, Murray RM, Elowitz MB (2008) Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics* **40:** 1493-1498.

Ehrt S, Guo XV, Hickey CM, Ryou M, Monteleone M, Riley LW, Schnappinger D (2005) Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Research* **33:** e21.

Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* **467:** 167-173.

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403:** 335-338.

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

Eydgahi H, Chen WW, Muhlich JL, Vitkup D, Tsitsiklis JN, Sorger PK (2013) Properties of cell death models calibrated and compared using Bayesian approaches. *Molecular Systems Biology* **9:** 644.

Flynn JM, Levchenko I, Seidel M, Wickner SH, Sauer RT, Baker TA (2001) Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 10584-10589.

Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, Gomes A, Rustad T, Dolganov G, Glotova I, Abeel T, Mahwinney C, Kennedy AD, Allard R, Brabant W, Krueger A, Jaini S, Honda B, Yu WH, Hickey MJ, Zucker J, Garay C, Weiner B, Sisk P, Stolte C, Winkler JK, Van de Peer Y, Iazzetti P, Camacho D, Dreyfuss J, Liu Y, Dorhoi A, Mollenkopf HJ, Drogaris P, Lamontagne J, Zhou Y, Piquenot J, Park ST, Raman S, Kaufmann SH, Mohney RP, Chelsky D, Moody DB, Sherman DR, Schoolnik GK (2013) The Mycobacterium tuberculosis regulatory network and hypoxia. *Nature*.

Geva-Zatorsky N, Dekel E, Batchelor E, Lahav G, Alon U (2010) Fourier analysis and systems identification of the p53 feedback loop. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 13550-13555.

Haario H, Laine M, Mira A, Saksman E (2006) DRAM: Efficient adaptive MCMC. *Statistics and Computing* **16:** 339-354.

Hasty J, Dolnik M, Rottschafer V, Collins JJ (2002) Synthetic gene network for entraining and amplifying cellular oscillations. *Physical Review Letters* **88:** 148101.

Hasty J, Isaacs F, Dolnik M, McMillen D, Collins JJ (2001) Designer gene networks: Towards fundamental cellular control. *Chaos* **11:** 207-220.

Hett EC, Rubin EJ (2008) Bacterial growth and cell division: a mycobacterial perspective. *Microbiology and Molecular Biology Reviews* **72:** 126-156, table of contents.

Izumo M, Sato TR, Straume M, Johnson CH (2006) Quantitative analyses of circadian gene expression in mammalian cell cultures. *PLOS Computational Biology* **2:** e136.

Joyce G, Williams KJ, Robb M, Noens E, Tizzano B, Shahrezaei V, Robertson BD (2012) Cell division site placement and asymmetric growth in mycobacteria. *PLoS One* **7:** e44582.

Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics* **6:** 451-464.

Karzai AW, Roche ED, Sauer RT (2000) The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nature Structural & Molecular Biology* **7:** 449-455.

Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, Frita R, Ten Bokum A, Besra GS, Lott JS, Stoker NG (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in Mycobacterium smegmatis and Mycobacterium tuberculosis. *Molecular Microbiology* **65:** 684-699.

Konopka T, Rooman M (2010) Gene expression model (in)validation by Fourier analysis. *BMC Systems Biology* **4:** 123.

Lin Y, Li Y, Crosson S, Dinner AR, Scherer NF (2012) Phase resetting reveals network dynamics underlying a bacterial cell cycle. *PLOS Computational Biology* **8:** e1002778.

Lipan O (2008) Systems biology. Enlightening Rhythms. *Science* **319:** 417-418.

Lipan O, Wong WH (2005) The use of oscillatory signals in the study of genetic networks. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 7063-7068.

Ljung L (1999) *System Identification: Theory for the User*: Prentice Hall.

Ljung L (2010) Perspectives on system identification. *Annual Reviews in Control* **34:** 12.

Locke JC, Elowitz MB (2009) Using movies to analyse gene circuit dynamics in single cells. *Nature Reviews Microbiology* **7:** 383-392.

Locke JC, Young JW, Fontes M, Hernandez Jimenez MJ, Elowitz MB (2011) Stochastic pulse regulation in bacterial stress response. *Science* **334:** 366-369.

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

Mettetal JT, Muzzey D, Gomez-Uribe C, van Oudenaarden A (2008) The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science* **319:** 482-484.

Mondragon-Palomino O, Danino T, Selimkhanov J, Tsimring L, Hasty J (2011) Entrainment of a population of synthetic genetic oscillators. *Science* **333:** 1315-1319.

Muttucumaru DG, Parish T (2004) The molecular biology of recombination in Mycobacteria: what do we know and how can we use it? *Current Issues in Molecular Biology* **6:** 145-157.

Pandey AK, Raman S, Proff R, Joshi S, Kang CM, Rubin EJ, Husson RN, Sassetti CM (2009) Nitrile-inducible gene expression in mycobacteria. *Tuberculosis (Edinb)* **89:** 12-16.

Park HD, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, Schoolnik GK, Sherman DR (2003) Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. *Molecular Microbiology* **48:** 833-843.

Russell DG (2007) Who puts the tubercle in tuberculosis? *Nature Reviews Microbiology* **5:** 39-47.

Rustad TR, Harrell MI, Liao R, Sherman DR (2008) The enduring hypoxic response of Mycobacterium tuberculosis. *PLoS One* **3:** e1502.

Rustad TR, Sherrid AM, Minch KJ, Sherman DR (2009) Hypoxia: a window into Mycobacterium tuberculosis latency. *Cellular Microbiology* **11:** 1151-1159.

Santi I, Dhar N, Bousbaine D, Wakamoto Y, McKinney JD (2013) Single-cell dynamics of the chromosome replication and cell division cycles in mycobacteria. *Nature Communications* **4:** 2470.

Sia SK, Whitesides GM (2003) Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies. *Electrophoresis* **24:** 3563-3576.

Singh A, Crossman DK, Mai D, Guidry L, Voskuil MI, Renfrow MB, Steyn AJ (2009) Mycobacterium tuberculosis WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. *PLOS Pathogens* **5:** e1000545.

Sliusarenko O, Heinritz J, Emonet T, Jacobs-Wagner C (2011) High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Molecular Microbiology* **80:** 612-627.

Smolen P, Baxter DA, Byrne JH (1998) Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. *American Journal of Physiology* **274:** C531-542.

Snapper SB, Melton RE, Mustafa S, Kieser T, Jacobs WR, Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Molecular Microbiology* **4:** 1911-1919.

Stover CK, de la Cruz VF, Fuerst TR, Burlein JE, Benson LA, Bennett LT, Bansal GP, Young JF, Lee MH, Hatfull GF, et al. (1991) New use of BCG for recombinant vaccines. *Nature* **351:** 456-460.

Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* **456:** 516-519.

Strogatz SH (2001) *Nonlinear Dynamics and Chaos*. Cambridge, Massachusetts: Westview Press.

Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440:** 545-550.

Suel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* **315:** 1716-1719.

Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 12795-12800.

Taylor RJ, Falconnet D, Niemisto A, Ramsey SA, Prinz S, Shmulevich I, Galitski T, Hansen CL (2009) Dynamic analysis of MAPK signaling using a high-throughput microfluidic single-cell imaging platform. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 3758-3763.

Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. *Nature* **457:** 309-312.

Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I (2006) The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. *Molecular Microbiology* **60:** 312-330.

WHO (2012) Global tuberculosis report 2012. WHO.

Young DB, Gideon HP, Wilkinson RJ (2009) Eliminating latent tuberculosis. *Trends in Microbiology* **17:** 183-188.

Young JW, Locke JC, Altinok A, Rosenfeld N, Bacarian T, Swain PS, Mjolsness E, Elowitz MB (2012) Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature Protocols* **7:** 80-88.

Curriculum Vitae