

Control Theory for Synthetic Biology

Recent advances in system characterization, control design, and controller implementation for synthetic biology

Victoria Hsiao*, Anandh Swaminathan*, and Richard M. Murray

*V. Hsiao and A. Swaminathan contributed equally to this work.

POC: A. Swaminathan (aswamina@caltech.edu)

November 14, 2017

Living organisms are differentiated by their genetic material – millions to billions of DNA
2 bases encoding thousands of genes. These genes are translated into a vast array of proteins,
many of which have functions that are still unknown. Previously, it was believed that simply
4 knowing the genetic sequence of an organism would be the key to unlocking all understanding.
However, as DNA sequencing technology has become affordable, even cheap, it has become
6 clear that living cells are governed by complex, multilayered networks of gene regulation that
cannot be deduced from sequence alone. Synthetic biology as a field might best be characterized
8 as a learn-by-building approach, in which scientists attempt to engineer molecular pathways that
do not exist in nature, and in doing so, test the limits of both natural and engineered organisms.

10 Synthetic biology broadly encompasses the genetic engineering of organisms in order to
implement and test new biological functions. A relatively young field, synthetic biology relies
12 on biological discoveries in gene function as well as improvements in molecular biology tools
for manipulation of DNA [1]. Current applications of synthetic biology include production of
14 biofuels and other valuable chemicals [2], [3], molecular computation and logic [4], [5], medical
diagnostics [6], and artificial microbial communities [7], [8]. These engineered biological circuits
16 are often not robust because of sensitivity to environmental conditions, context effects within the
host organism, and stochastic noise due to inherently low molecular counts. Applying feedback
18 control would potentially allow biological circuits to perform their intended function more
robustly across a variety of operating conditions, and ease the transition from very controlled
20 lab conditions to practical real world applications.

This survey aims to provide a general overview of relevant terms and resources for
22 understanding the intersection of synthetic biology and control theory. A reader with a
background in control theory should come away with a reasonable understanding of the current
24 state-of-the-art of biological system identification, controller design and implementation, and

the open challenges facing the field. Additionally, this review updates and builds upon previous publications on this subject [9], [10], [11]. As this particular work is limited to a selected number of topics, additional reviews are suggested throughout the text for deeper reading.

In the following sections, each of the challenges is addressed within the typical workflow for control implementation of more traditionally engineered systems (Figure 1). Engineered biological systems present a number of challenges to all stages of this workflow for reasons such as limitations in real-time measurement, resource competition with the host organism, and incomplete knowledge of underlying biological processes. First, strategies for framing a biological organism as a system with defined inputs, outputs, sensors, actuators, and measurements are discussed (Figure 1a). Obtaining dynamic and reliable measurements within biological organisms is a daunting challenge, engineered or otherwise. An overview of the state-of-the-art tools for modeling and characterizing biological systems is presented, followed by system identification methods specifically designed for the types of data available from biological measurements. The difficulty in engineering complex genetic networks, combined with severe limitations in real-time measurement, means that the body of work for controller design (Figure 1b) is limited – as a result, we discuss the open problems and challenges awaiting the entrepreneurial reader, and also present a number of examples of feedback loop implementation in living cells (Figure 1c). Finally, the necessary challenges in synthetic biology and development of control theoretical frameworks that need to be addressed in order to advance the field are discussed.

Introduction to Synthetic Biology

In the last two decades, advances in molecular biology techniques have led to the development of reliable methods for extraction, duplication, and insertion of DNA elements into non-native host organisms. Combined with the advent of affordable commercial DNA sequencing, these cloning techniques paved the way for the beginnings of synthetic biology. The universality of the central dogma (i.e. all organisms store information in the form of DNA) means that genes previously identified from many different types of prokaryotic and eukaryotic organisms can be inserted and expressed in a non-native host organism (Figure 2a; see also “Sidebar: Central Dogma”). Host organisms are usually so-called “model organisms” that have been domesticated to increase efficiency of taking on foreign DNA and growth under controlled laboratory conditions. The most commonly used model organisms include the gram-negative bacteria *Escherichia coli*, the gram-positive bacteria *Bacillus subtilis*, the eukaryotic yeast *Saccharomyces cerevisiae*, and the mammalian Chinese Hamster Ovary cell line (CHO). Synthetic biology takes the approach of an engineering discipline, albeit one in which many of the moving parts are not well characterized nor optimally designed for the function that is

being built. While there is a large emphasis on building molecular systems *de novo*, synthetic biology relies on protein characterization and discovery from biology and biochemistry, as well as continued elucidation of natural molecular networks from systems biology.

The potential applications for safe and reliably engineered organisms for future societies are vast. Biofuels could be produced sustainably and on-demand in programmable fermentation tanks. Medicines from rare or endangered plants could be produced in distributed fermentors, reducing production costs and reliance on weather and rainfall. Nutritional deficiencies could be corrected with engineered gut bacteria that can detect and produce essential vitamins or break down accumulated toxins. Organisms for bioremediation could be tailored to specifically respond to the type of environmental contamination. While many of these concepts have been proven in the laboratory, the field faces major challenges in reliability and robustness outside of extremely controlled conditions. Inherently, synthetic biomolecular systems must compete with natural processes for survival and replication, resulting in a strong evolutionary pressure for cells to turn off or disrupt engineered pathways.

Published works in synthetic biology can be loosely separated into two categories: modular circuits to demonstrate and characterize the dynamics of novel biomolecular networks (synthetic gene circuits), and the expression of large metabolic pathways in new host organisms for the purposes of large scale production of a specific product (metabolic engineering). Figure 2b illustrates several well known synthetic circuits. One of the first, and most widely known, examples of a synthetic circuit is the *repressilator*, named as such for its oscillating output behavior generated by a network of three transcriptional repressors [12]. In the same year, the bistable toggle-switch utilized mutual repression of two proteins to create a memory switch [13]. Since the publication of these studies in 2000, there have been many additional examples of oscillators [14], [15], [16], [17], switches [18], logic gates [19], [20], [21], [22], [23], counters [24], finite state machines [25], [26], and memory [27], [28], [29]. Quorum sensing systems from different microbes can effectively be used for artificial cell-cell signaling for predator-prey systems [30] and two-dimensional pattern generation [31].

In parallel with these completely synthetic gene circuits, there has been a huge effort to express entire transplanted pathways for the large-scale production of small molecules. In metabolic engineering, the engineered host organisms are used as catalysts to convert a simple carbon source, usually glucose or other sugar, into complex chemical and protein products (Figure 2c). Notably, this has led to the commercial scale production of molecules such as the anti-malarial drug artemisinin [32], the biofuels farnesene [33] and 1, 4-butanediol [34], and polymers such as spider silk [35]. With industrial synthetic biology, chemicals and materials that have traditionally been impossible to farm at large scale, or harvested from increasingly endangered

plants, can be produced at commercial scale in a renewable way.

2 Many challenges exist in improving robustness of engineered strains. At the heart of the
issue is that synthetic protein production competes for resources with existing cellular processes
4 for survival and replication (Figure 2d). Synthetic gene circuits usually rely on the heavy
overexpression of non-native proteins, an energy-intensive process that diverts away from the
6 common pool of cellular resources. Furthermore, even if the synthetic pathway is regulated
with a switch, these molecular mechanisms can be “leaky”, resulting in unwanted background
8 expression even when “off”. In [36], the relative carrying capacity of engineered *E. coli* was
measured by using growth rate as a proxy for carrying capacity. In this study, the growth of cells
10 with an active synthetic circuit was compared against the growth rate of cells with an empty
plasmid vector. This data was then used to determine designs to minimize load on the cell. At
12 the first possible opportunity, host organisms will find ways to mutate away or repress synthetic
pathways that are particularly deleterious to growth [37]. As a result, synthetic biologists have
14 relied on the use of antibiotic selection, nutrient auxotrophy, or toxin-antitoxin systems to select
against cells that mutate away from the original design [38]. While these strategies are not
16 covered in detail here, a review on building in biosafety can be found in [39]. The ideal design
would force the host to rely on the synthetic circuit, and for the synthetic circuit to only be
18 functional in that specific host. However, without careful design, these strategies could add to
the cellular cost of implementing the synthetic circuit and may not be reliable at scale.

20 Concurrent with building experimental systems, there has been a large body of work on
modeling these synthetic circuits, particularly since the complexity of the cell requires a great
22 deal of assumptions and curation to decide which parts of the system need to be represented.
Model classes (see: “Sidebar: Model Classes”) and implementation will be discussed in detail
24 in later sections. Due to the limited scope of this review, we will focus exclusively on studies
on either engineered or natural biological systems that feature feedback control.

26 **Natural Inspiration for Controller Design**

Endogenous systems in organisms display many principles of feedback control and can
28 be sources of inspiration when designing controllers. These networks display a wide array of
performance characteristics and robustness to perturbations that engineered *de novo* biomolecular
30 networks have yet to achieve.

In general, integral control is crucial for natural biomolecular pathways because organisms
32 must maintain tight regulation over all aspects of their internal environment (*e.g.* temperature,
pH, nutrients), a phenomenon known as homeostasis. The zero steady state error property

makes integral feedback an excellent choice for robust regulation of the cellular environment.

In cows, the concentration of calcium in the blood is maintained when the demand for calcium increases by a control system involving two hormones: parathyroid hormone (PTH), and 1-25-dihydroxycholecalciferol (1,25-DHCC) (Figure 3a). It is shown in [40] that PTH implements a proportional response to a calcium shortage while 1,25-DHCC implements an integral response over a longer timescale. Together, the two hormone system makes a PI controller, where a fast proportional term removes the majority of the error quickly, and a slow integral term leads to eventual zero steady state error. Other examples of integral control in natural systems that are not discussed in detail here include bacterial chemotaxis [41], [42] and yeast osmoregulation [43]. Finally, in [44], there is a comprehensive summary of the different synthetic gene circuit topologies that are known to be capable of providing near exact adaptation.

Recently, positive feedback was proposed as a mechanism for maintenance of bacterial persisters within populations [45] (Figure 3b). Compared to the rest of the population, persisters have greatly reduced nutrient intake and growth, which actually endows higher levels of antibiotic tolerance, since many antibiotics target cellular machinery for replication and metabolism. This can result in survivors that can mutate and develop resistance. The proposed mechanism of positive feedback is that stochastic entry into a low metabolic flux state becomes perpetuated by internal stress responses. Cells in the persister state can only be rescued back into the high metabolic flux state with stochastic increases in metabolic enzymes or stochastic decreases in production of growth-inhibiting proteins. There are also quite a few examples of positive feedback as a mechanism for trigger waves. Trigger waves, unlike diffusing molecules, do not slow down nor lose amplitude as they propagate. Examples of trigger waves include neuronal action potentials, calcium waves across tissues, and mitotic waves in *Xenopus* frog eggs [46].

Other studies have attempted to elucidate the complex interactions of natural networks. In [47], some of the multiple feedback pathways that regulate galactose metabolism in the yeast *S. cerevisiae* were investigated (Figure 3c). Galactose is one of many sugars that yeast can metabolize, and in the absence of galactose the pathway is kept in the OFF-state. This is achieved by a mechanism in which a repressor protein, Gal80, actively sequesters an activator protein, Gal4. In the presence of galactose, Gal80 is itself sequestered by two additional proteins, Gal3 and Gal1. Gal3 and Gal1 independently amplify their own production with positive feedback. Through a combination of experimental knockouts and simulation, [47] demonstrates that this positive feedback is essential for creating a bimodal response to galactose in yeast.

Studies on the regulation of single cell competence in the gram-positive bacteria *B. subtilis* suggest similarly nested positive and negative feedback loops (Figure 3d) [48]. Competence is a cellular condition in which a single cell takes in foreign DNA from the environment, and

is triggered by stress response. In [48], it is found that activation of bacterial competence is regulated through competitive inhibition of the protein degradation machinery. The regulatory network is characterized by fast positive feedback for the main regulator ComK, but slow negative feedback by ComS. ComS competitively inhibits degradation of ComK, allowing ComK to activate the competence genes. ComS is up-regulated by the stress response genes and so its buildup results in accumulation of ComK. However, ComK also indirectly inhibits ComS, thus allowing the cell to turn off competence and return to the normal vegetative state.

As the mechanisms of natural signaling pathways become better understood and more accurately measured, it is clear that biology takes full advantage of redundancy, fast and slow time responses, molecular sequestration and titration, and nested feedback to maintain near-perfect adaptation and response to the constant flow of perturbations. Synthetic biology has just only begun to imitate these natural networks.

Characterization and Identification of Biological Systems

System identification is particularly challenging for biological systems because a majority of the processes that occur within the organism are part of the underlying host biology, which may not be well understood and also difficult to measure. This can result in severe limitations on creating accurate models of synthetic gene circuits that can be used for control design or other computational explorations of circuit behavior (see: “Sidebar: Uncertainty in Biology”). Considerations for model choice and methodology for system identification are largely dependent on which types of data can be collected (see “Sidebar: Model Classes” and “Sidebar: Measurement Types”). This section will feature discussion of results for identifiability of synthetic gene circuits, followed by a description of methods for deterministic identification from bulk measurements. Finally, an overview of techniques for inferring parameters for stochastic biological circuits from single cell population snapshots or time series data will be discussed. Figure 4 illustrates the procedure of Bayesian system identification for both deterministic and stochastic synthetic gene circuits.

Identifiability of Synthetic Gene Circuit Models

Theoretical and computational considerations of the identifiability of synthetic gene circuits tend to focus on whether circuit parameters are identifiable for stochastic models given single cell population snapshot data. This is because these types of data are not typically collected in other engineering disciplines. However, even for bulk data and deterministic models, there are specific issues with system identification for synthetic gene circuits.

As the identifiability problem is difficult to assess, theoretical results tend to focus on linear systems where the theory is tractable. In [49], the observability problem for linear systems in which the linear system evolves a distribution over time was considered. Within this framework, the linear system represents the dynamics of a synthetic gene circuit, and the distribution at each time represents the distribution of the state of the circuit across a large population of cells. The output is then the distribution of an output for all time. The observability problem is to reconstruct the initial distribution of the system given the output distribution at every time. This distributional observability problem was shown in [49] to be strictly harder than the standard observability problem of finding a single initial state from a single output trajectory. Furthermore, it was proven that a sufficient condition for reconstructing the initial state distribution is to have an observable system with an output dimension of at least $n - 1$, where n is the state dimension. Finally, the problem of structural identifiability of linear systems driven by white noise is considered in [50]. In this system, the entire steady state covariance of the system can be measured over time, but only certain subsets of state can be measured in time series. It is assumed that even though the system itself is stochastic, measurements are perfect and the steady state covariance and output correlations can be measured exactly. Results from [50] showed that a combination of the steady state covariance and output correlations can be sufficient to determine the dynamics of the system exactly. This work is relevant to biological circuits, because often many species can be measured simultaneously in steady state population snapshot measurements, but only a few species can be measured simultaneously in time series measurements.

Computational approaches to the identifiability problem rely on using simulated data sets to test the performance of system identification. In [51], the problem of identifying the parameters of a stochastic model of a genetic toggle switch given bulk data on the circuit, marginal distributions of the species, or the full joint distribution of the species in the system is addressed. The results show that only the full joint distribution is sufficient to recover uniquely the true parameters of the model.

Identification of Deterministic Biological Circuits

Techniques for identification of deterministic models of biological circuits are often quite similar to the techniques used for general identification of nonlinear systems. However, for synthetic gene circuits, data can typically be collected only for a few outputs at limited time resolution. Limited time resolution makes it hard to filter the time derivative of the output, which makes it difficult to apply a lot of traditional system identification methods that rely on the derivative of the measured output. Furthermore, most biological models contain many species, reactions, and parameters, and because only a small number of outputs can be measured, there are often many sets of parameters that can fit the measured data equally well. This effect,

known as “parameter degeneracy,” can make it difficult to resolve the best parameters for the modeled system. Many techniques for system identification for synthetic gene circuits focus on alleviating the issue of parameter degeneracy.

For example, the extended Kalman filter is often used for parameter estimation, where the state is augmented with the parameters. However, [52] demonstrated that in a biological setting, matching the noise characteristics of the filter in addition to fitting the trajectory to measured data can help resolve parameter degeneracy. In [53], [54], relaxation type procedures are used to consider parameters sets that are consistent with the measured data, resulting in a set of regions in parameter space that fit the data within some error tolerance. Additionally, standard techniques based on Markov chain Monte Carlo (MCMC) can also be used [55]. Another approach to system identification for nonlinear systems is to build a dictionary matrix based on candidate functions and then use ℓ_1 regularization to select only a subset of the candidate functions to model the dynamics [56]. However, this approach requires full state measurements, which limits its applicability to synthetic gene circuits.

Identification for Stochastic Biological Circuits from Population Snapshots

With population snapshots, a large amount of data can be collected for a group of cells for a single moment in time. While a population can be continuously sampled and measured, each sampling will be a different group of cells. Generally, population snapshots are not resolved in time, but can generate more data per sample than a time-course experiment (See: “Sidebar: Measurement Types”). Identification methods for stochastic biological circuits from population snapshot data can be categorized into three different types: methods based on optimization and the finite state projection, methods based on moment closure, and Bayesian methods based on stochastic simulations of the system and MCMC.

In the optimization-based methods, the finite state projection method [57] is used to create a large linear system whose state is approximately the distribution of the chemical species in the system. The state distribution time trajectory can be calculated by solving the linear system, and the likelihood function can be computed from the observed data using a multinomial likelihood function. This method has been used to discriminate between different models of yeast osmotic stress response [58]. In [51] and [59], the same method is applied, but instead of a multinomial likelihood, a ℓ_1 norm penalty on the difference between the observed and predicted distribution is used. These types of methods work well for systems with small state spaces because the full state distribution can be solved. However, for systems with many species or high molecular counts, the finite state projection yields increasingly huge linear systems that become prohibitively expensive to repeatedly evaluate. Note that while [57], [59], [51], [58] all use the finite state projection

to do an optimization based fit of the parameters in their models, the same likelihood function could be used in a Bayesian inference scheme as well.

Moment closure methods are one way to get around the curse of dimensionality in modeling stochastic gene circuits. With moment closure methods, the moments of the system can be modeled over time using a system of ODE's and these simulated moments can be compared with the experimentally measured moments for different parameter sets. Though these methods are known to struggle in instances where the distributions are bimodal, they have been shown to successfully predict an observed bimodal experimental data set in limited cases [60]. In other cases, moment based methods can be used to characterize the causal relationships between different genes [61]. In [62], moment based inference in conjunction with an optimal experiment design technique based on maximizing the expected Fisher information is used to characterize a light inducible synthetic gene circuit. The limitation of moment closure methods is that the equations for the N th order moments usually depend on higher order moments, and so to avoid having an infinite dimensional system of ODE's, the higher order moments have to be assumed to have some known form. In some cases, the higher order moments can be computed from the lower order moments if the distribution is of a known type such as a normal or log-normal distribution [63]. In other cases, the third order cumulants are assumed to be zero [60]. A data-based identification procedure based on moment closure has also been developed, in which the reaction rate parameters are identified from the mean moment equations and the covariance values are plugged in from experimentally measured data [64]. In this way, no prior assumptions on the distributions of the chemical species are required. Finally, semidefinite programming can be used to compute upper and lower bounds on the moments of a stochastic differential equation [65]. This approach is more computationally expensive than simple moment closure but provides a theoretical guarantee that the moment of interest lies within a specific interval.

The third class of methods for identifying circuits from population snapshot data is methods based on repeated stochastic simulations of the system. Some methods are based on approximate Bayesian computation [66], where the data is simulated for many different parameter sets and parameter sets that produce results in agreement with experimental data are kept. The INSIGHT method developed in [67] uses concentration inequalities to bound the deviation between the experimentally observed cumulative distribution function (CDF) of a certain output and the simulated CDF for a given parameter set. This deviation decreases as the sample size increases. With this technique, the method is able to quickly reject bad parameter sets after a very small number of simulations, vastly improving computation speed.

While stochasticity is one way to explain the variation in circuit output across a population of cells, another possible explanation is heterogeneity between cells. Not every cell has exactly

the same environment, and if different cells have different parameters and initial conditions, then the output will follow a distribution even if the circuit dynamics within each cell are deterministic. In [68] and [69], there is both the assumption that the circuit dynamics are deterministic, and that parameters for the cells are sampled from an underlying distribution or that there is a discrete number of cell types. This leads to a Bayesian inference procedure that can recover the parameters for each cell type, or the parameter density for a heterogenous population of cells.

Identification for Stochastic Circuits from Time Series Data

Single cell time series data usually involves using microscopy to track the dynamics of a small sample of single cells for an extended period of time (See: “Sidebar: Measurement Types”). While these data are limited in sample size, the time resolution can be on the order of milliseconds over a period of minutes to hours. Identification of stochastic biological circuits from single cell time series data is typically done using methods based on Approximate Bayesian Computation or particle filtering. By selecting parameter sets and simulating a data set for each set of parameters, it is possible to perform Bayesian inference using MCMC without having to explicitly compute likelihoods [70]. This method is used in [55] to perform system identification on a simulated data set generated by a stochastic model of a repressilator. This type of likelihood-free method does not work well when the data is of a large dimension, and a better method is the particle marginal Metropolis Hastings method, where a particle filter is used to approximately compute the likelihood of the data given a set of parameters [71]. In [71], methods for speeding up the inference by using diffusion approximations to the stochastic simulation procedure are proposed.

Heterogeneity between cells can also be an issue for system identification. In [72], the heterogeneity between cells is modeled by assuming each cell has extrinsic factors and rate parameters distributed according to a gamma distribution. This choice of distribution allows for integrating out the extrinsic factors and parameter uncertainty while performing Bayesian inference. This approach was termed ‘dynamic prior propagation’, and was demonstrated on experimental data obtained from an inducible synthetic circuit in yeast.

Biological Controller Design

There are two main ways to design controllers for synthetic gene circuits. The first way is to design the controllers to be synthetic gene circuits in single cells. In this case, each cell has its own copy of the controller. However, because the controller must be implemented using a set of chemical species and reactions, the controller structure is highly constrained. Additionally, because the controller species themselves are subject to stochastic fluctuations,

there are fundamental limits on how well the controller can control the downstream process.

2 To alleviate these issues, controllers can also be implemented at the population level using a
computer in the loop. In this case, the controller dynamics can be arbitrary, but the control must
4 be implemented at the population level, as current technology does not allow for independent
actuation of distinct single cells. This section will discuss the fundamental limits for control of
6 synthetic gene circuits, followed by controller design for synthetic gene circuits in single cells
or with a computer in the loop.

8 **Fundamental Limits of Control**

While the fundamental limits of controller performance associated with standard control
10 design still persist in synthetic gene circuits, stochasticity introduces a new class of hard limits
on performance. In traditional control design, Bode's integral formula [73] provides a hard limit
12 on controller performance.

Stated formally, if a system has a loop transfer function $L(i\omega)$, then the closed loop transfer
14 function from disturbance to output is given by the sensitivity function, which is another transfer
function given by

$$S(i\omega) = \frac{1}{1 + L(i\omega)} .$$

16 It is desirable to make the magnitude of sensitivity function small in order to minimize the
effect of disturbances on the output. However, for a loop transfer function that has no right half
18 plane poles and at least two more poles than zeros, Bode's integral formula gives the following
constraint.

$$\int_0^{\infty} \ln |S(i\omega)| d\omega = 0 \quad \text{(Bode's integral formula)}$$

20 Therefore, in order to mitigate the effects of disturbances and reduce the sensitivity function
at one frequency, it is necessary that the sensitivity function increases and disturbance rejection
22 worsens at another frequency. This effect is informally termed the waterbed effect, because
pushing down on the sensitivity function at one frequency causes it to increase at another
24 frequency. Effects reminiscent of the waterbed effect have been observed in natural biological
systems.

26 For example, in [74], yeast are subjected to an environment that oscillates between normal
salt levels and high salt levels that induce osmotic stress. During the course of the experiment, the
28 yeast spend half of the total time in the stress environment. As the frequency of the environmental

switching was varied from every 30 seconds to every 128 minutes, it was observed that at the fastest and slowest time scales, the yeast grew at a rate similar to an environment with no stress. This suggested that the yeast could adapt very well to an environment that fluctuates very slowly or very quickly. However, when the high salt condition was switched on and off at 8 minute intervals, the yeast grew much more slowly. The stress response performance is thus tuned for excellent performance at very high and low frequencies at the expense of performance at intermediate frequencies, which is consistent with the waterbed effect. Whether or not this limitation occurs because of mechanistic hard limits in the yeast's control architecture is unknown, but the results are consistent with the limits on control imposed by Bode's integral formula.

In [75], control theory was applied to gain additional understanding of a phenomenon known as glycolytic oscillations, in which ATP levels oscillate in yeast. Adenosine triphosphate (ATP) is the molecule unit of energy within all living cells. The oscillations were explained as the result of fundamental trade-offs in a control system designed to control ATP levels in the cell. An essential feature of robust control theory is the trade-off between robustness and performance of a control system [73]. In [75], because ATP is the source of energy in cells, the ATP supply control system is tuned with a heavy bias towards performance, which means that it is less robust and prone to losing stability, thus creating glycolytic oscillations.

In addition to the traditional hard limits on control performance, stochasticity introduces further limitations. Because the reactions and species that implement the controller are subject to stochastic fluctuations, the controller, which is supposed to reduce noise in the system, can also introduce additional noise. For a general class of feedback systems, [76] showed that the standard deviation of the output distribution goes as the inverse fourth root of the number of signaling events used by the controller. That is, if the goal is to decrease the standard deviation of the output by 10-fold, then 10,000 times more signaling is needed. The resulting trade off is then one between output regulation and signaling effort, which is common in standard optimal control. However, certain classes of biological systems fall outside the scope of the systems considered in [76]; these classes of systems may not be subject to such devastating fundamental limits.

Control Design in Single Cells Using Gene Circuits

In controllers implemented in single cells with gene circuits, it is difficult not only to have general controller structures but also to quantitatively tune gains. Because the controller must be implemented in cells with a set of chemical reactions, the chosen set of chemical reactions sets the structure of the controller. Tuning the gains for the controller then consists of selecting

different genetic parts that can modulate the expression of the controller chemical species. This is typically done by varying the ribosome binding site (RBS) sequences for the controller species using a library of RBS's with known relative strengths [77]. However, the absolute strengths of these parts depend on context and are typically not known. Therefore, instead of setting one absolute optimal gain, the control designer must do multiple iterations to tune the gain in a relative manner. A comprehensive review of different types of genetic parts and their respective tunability can be found in [78].

The control design then consists of two phases. The first phase is one of selecting the synthetic genetic parts that will implement feedback, and the second phase consists of tuning the gains to achieve the desired control metrics. Mathematically, the first phase of design effectively sets the model of the closed system, where a set of ODE's or stochastic chemical reactions describes the process and controller reactions. The second phase then consists of optimizing the tunable parameters in the model to get the desired closed loop behavior.

With respect to the first phase, many researchers have worked on how to design chemical reaction systems that implement integral control. In [79], [80], a general model of integral control was studied and it was shown that in order to have integral control, one of the ODE's in the system should depend effectively only on the output. For example, the rate of change of a control species x might look like

$$\frac{dx}{dt} = f(y),$$

where f is some function and y is the output that needs to be regulated. Then, if the system reaches a steady state, y must satisfy $f(y) = 0$. If f is selected such that its only root occurs at the desired steady state value of y , then if the system reaches an equilibrium, the output y is guaranteed to have the desired steady state value.

In practice, dilution due to cell growth can cause the dynamics of x to depend on the amount of x present. However, if the controller acts on a much faster time scale than cell growth, the dilution can be ignored. This system implements a general version of integral feedback, where x in some sense integrates the output over time and ensures that y always has the same steady state value.

In [81], it is shown that exact adaptation can be accomplished without even having negative feedback in metabolic flux networks. A minimal example of a reaction system discussed is one where A is produced with a constant rate, metabolized to B , and B is then degraded. The

dynamics of this system are given by the equations

$$\dot{A} = k_0 - k_1 A, \tag{1}$$

$$\dot{B} = k_1 A - k_2 B. \tag{2}$$

It is then straightforward to verify that the steady state value of B of the system exhibits perfect adaptation with respect to variation in the parameter k_1 .

In the stochastic regime, [82] developed a theory of stochastic integral control, where an arbitrary reaction network can be augmented with 4 simple reactions to accomplish mean reference tracking of an output species. By using time evolution equations for the moments of the species in the system, it was shown that the first moment of the output species must converge to the desired reference value if the system converges to a steady state. In this case, the steady state of the system is a distribution, and thus ensuring convergence consists of showing that the system of stochastic chemical reactions is ergodic. This result holds for a large class of stochastic systems, and its simplicity is promising for implementation in single cells.

The main limitation of integral control in cells is the fact that in growing cells, all species are constantly subject to dilution due to cell growth. In both the deterministic and stochastic case, dilution had to be ignored in order to get zero steady state error. In actual implementation, integral control should perform well as long as the actual reactions in the system and controller occur on a much faster time scale than dilution, thus minimizing dilution's effect on integral control performance. The time scale of dilution varies according to the host organism; *E. coli* have a doubling time of 20 minutes, while mammalian cells can take hours to days. Figure 5 illustrates examples of biological integral control in both the deterministic and stochastic regimes. For a more detailed explanation of the examples in Figure 5, please see "Sidebar: Integral Control for Synthetic Gene Circuits."

In addition to the theoretical work on picking optimal control structures for integral control, much work has also been done on analyzing and optimizing parameters for a given control architecture. In [83], a mathematical model was used to explore the possibility of using feedback control to improve biofuel production in *E. coli* (Figure 6). Biofuels are toxic to *E. coli*, and so efflux pumps must be produced to export the biofuels out of the cell (Figure 6a). However, producing too many efflux pumps is toxic to the cells as well. Therefore, using feedback to produce efflux pumps only when too much biofuel has accumulated could improve overall biofuel production. The authors consider a number of different biologically reasonable implementations of feedback – no feedback (constant production of pump proteins), an open loop response to biofuel levels, a repressor cascade, and a feedforward loop (Figure 6b-e). Each of these implementations has a unique associated ODE model. For each model, the authors

optimize biofuel production by tuning the model parameters. As expected, they find that the performance of a feedback control strategy is more robust to unknown parameter variation than an open loop strategy. Notably, the results showed that a loop combining negative feedback with a feed forward system had the optimal performance (Figure 6f). In this system, excess intracellular biofuel induces a small immediate increase in efflux pump production followed by a delayed large increase in pump production if the biofuel level stays high. This helps filter out sporadic short term biofuel accumulation due to intrinsic variation in the system, while still responding appropriately to a long term build-up of intracellular biofuel.

An important consideration for synthetic gene circuits is the limited amount of resources available within in each cell for the expression of circuit genes. In [84], modeling is used to guide the design and testing of an activation cascade gene circuit. Each gene in the cascade is supposed to activate the next gene, but in fact, a gene can also repress a downstream gene by competing with it for resources. In [84], modeling and analysis is used to design an activation cascade that actually preserves full activation. A similar type of work that considers feedback synthetic circuits will likely be necessary in order to successfully implement complex controllers in cells.

Balancing cellular resources and understanding relative kinetics are crucial for identifying potential bottlenecks. In [85], a system using proportional feedback to control gene expression was modeled. Detailed parametric analysis of the model found limitations due to saturation in the rates of different component reactions and the limited dynamic range of other processes. In [86], regulation of metabolic pathways was investigated by tuning enzyme production and feedback to avoid failure cases in which an excess of pathway intermediates or products build up. This work was extended in [87] to consider noise propagation in a metabolic pathway with feedback. In the case of a low-yield metabolic pathway, it is shown that a synthetic feedback circuit may amplify the stochastic noise of the pathway. Design strategies are suggested for attenuating this noise propagation.

Multicellular systems and control theory

Tools from control theory have had an especially notable impact in the design of multicellular synthetic gene circuits. In these circuits, interactions between cells play a role in addition to the typical genetic interactions that occur within single cells. Here, we describe three representative pieces of research in which control techniques were used to aid in the design or analysis of multicellular systems.

First, [88] showed that multicellular gene circuits can improve the robustness of the genetic

toggle switch [13]. As shown in Figure S5 (“Sidebar: Model Classes”), the standard toggle switch will spontaneously switch steady state due to noise after some period of time. A new toggle switch design was proposed, one that operates at the population level in which cells communicate with each other and influence each other’s steady state. In this case, even if a few cells switch their steady state due to noise, the rest of the population can drive those cells back to the correct steady state. Thus, the communication between cells makes this multicellular toggle much less likely to randomly switch steady states. Notably, the design in [88] relies heavily on monotone systems theory [89] to show theoretically that spontaneous errors in single cell steady states will be suppressed.

On the analysis side, stability analysis techniques have been used to determine the stability of gene circuit dynamics in networks of cells [90]. In this case, each cell is modeled as having its own system of nonlinear ODE’s describing its internal dynamics, and the interactions between cells are modeled as occurring on a graph with an associated interconnection matrix. Theoretical conditions for network stability are derived. These conditions depend on both the internal cellular dynamics as well as the interconnection structure. These types of techniques could be applied to predict the behavior of cells laid out spatially, where each cell can only communicate with neighboring cells.

As a third example, [91] implemented a synthetic gene circuit design for controlling the population size of a microbial culture. As the engineered population increased in size, the cells produce more of a lethal toxin, creating a negative feedback loop that leads to increased cell death. Recently, work in [92] demonstrates that the controller designed and implemented in [91] is a lag compensator, which is a commonly used controller for reducing steady state error.

Control Using a Computer in the Loop

Another way to get around the issues of controller implementation in single cells is to implement a controller using a computer in the loop. In this case, the cells are controlled at the population level. Using an experimental setup, the output (usually florescence) is measured for a cell population, and then a computer running the control algorithm decides the optimal input, which is then fed back into the system. The details of the implementation are covered in the next section. Here, we briefly review the control design for these systems.

As is the case with many other complicated nonlinear systems, the control scheme of choice for controlling synthetic gene circuits with a computer in the loop is model predictive control (MPC). In MPC, a model of the system is used to plan a trajectory and associated set of control inputs for a certain number of time steps. However, only one of these steps is actually

executed, and then the procedure is repeated [93]. The control scheme then consists of two parts: a state estimator and a controller. For synthetic gene circuits, the controller is typically a standard model predictive control based controller, where the optimal control input is chosen by solving a non-convex optimization problem at each time step over all possible input sequences. The state estimator is usually based on either an extended Kalman Filter (EKF) or a particle filter.

The studies described in [94] and [95] both use similar control schemes of combining an extended Kalman filter with a model predictive controller to control gene expression in a population of cells. The model predictive control problem is then solved with general purpose optimization software. In [96], a particle filter is used in place of an extended Kalman filter for state estimation in order to better account for uncertainty in the state estimate. In this case, the model predictive control input is optimized across a distribution of unknown initial states. This is again solved with general purpose optimization software. Notably, it was found that the combination of a particle filter for state estimation and model predictive control performs better than a simple PI controller with tuned gains. In [97], PI control of gene expression was performed. However, the PI controller was combined with a predictive scheme to help account for delays in the actuation of the system. A worked example of model predictive control using a toy model based on [95] can be found in “Sidebar: Model Predictive Control of Gene Expression.”

Implementation of Synthetic Biological Controllers

The implementation of synthetic biological controllers involves considerations for the timescales of molecular biology and gene expression, the modularity and availability of genetic components, and the types of inputs and outputs that can be applied to an engineered organism in real-time. This section covers these topics, along with examples of implemented biological controllers. Additional information on the molecular biology workflow and biological measurement types can be found in “Sidebar: A Molecular Biology Workflow” and “Sidebar: Measurement Types.”

Molecular libraries for implementation of biological controllers

One main motivation for characterization of biological systems is to create libraries of synthetic genetic parts that can be used when assembling more complex circuits. Two common issues that arise when developing libraries of genetic parts are orthogonality and context dependence. In most instances, genetic parts are not modular and their performance depends heavily on their environment. In addition, different parts can suffer from crosstalk with one another, which may create unwanted interactions that break circuit behavior. The idea behind well

characterized part libraries is to create sets of orthogonal parts that have predictable performance
2 in different contexts. These parts would then be useful for assembling larger synthetic circuits.

The basic ‘unit’ of a synthetic gene circuit consists of a promoter, a ribosomal binding site
4 (RBS), and a coding sequence for the protein of interest (Figure 7a). Each of these elements has
been characterized and engineered in attempts to regulate and predict the behavior *in vivo*. The
6 promoter regulates how frequently mRNA gets transcribed from the DNA sequence. The RBS
regulates how frequently the mRNA gets translated into a chain of amino acids, which then fold
8 into a complex three-dimensional structure known as the protein.

Transcription is the process in which RNA polymerase unwinds the DNA and encodes
10 the information in messenger RNA (mRNA). Transcriptional regulation has historically been
achieved through a combination of engineering of the DNA promoter sequence itself to affect
12 RNA polymerase binding efficiency, and through the use of additional protein repressors or
activators that either obstruct or recruit the polymerase (Figure 7a). The most commonly used
14 repressors in synthetic biology are known as TetR and LacI, while the most commonly used
activators are AraC and LuxR. While these commonly used parts have specific binding sequences
16 that must be incorporated into the promoter, any sequence of DNA can now be targeted with
activator or repressor variants of the dCas9 protein [98]. Protein activation or repression is
18 usually characterized in terms of fold change between the OFF state and the ON state. Once it
can access the promoter, RNA polymerase typically transcribes at a speed of 10–100 DNA bases
20 per second, with genes varying from ~ 1 kilobase in bacteria to 10 kilobases in humans [99]
(Figure 7b). Most of the promoters for each host organism were found because they regulate
22 important processes within the host. In order to be completely orthogonal to host functions, much
work has been done to engineer additional combinatorial regulation [100], developing orthogonal
24 repressors [101], and engineering completely synthetic promoters [77].

Ribosomal binding sites (RBSs) dictate the number of times a single mRNA gets translated
26 before it is degraded; Ribosomes translate mRNAs at rates on the order of $\sim 3 - 10$ amino acids
per second [99] (Figure 7c). While RBSs have been relatively well characterized for translational
28 efficiency and final protein production levels [102], it has also been observed that RBS expression
strength varies relative to the downstream sequence of the coding region [103]. To investigate
30 and address the problem of context dependence, a set of ribosome binding sites was engineered
to have similar relative strengths in gene expression independent of the adjacent promoter and
32 gene [77]. Since a large part in current engineering of gene circuits involves fine-tuning of gene
expression levels, any tool that can alleviate context issues becomes immensely valuable.

34 The half-life of a protein can be tuned by the addition of degradation tags (Figure 7d),

which are short amino acid sequences recognized by housekeeping proteases (enzymes with
2 the specific function of breaking down other proteins). Unlike mRNA, proteins are usually
remarkably stable in biological conditions. Without a degradation tag, it is generally assumed
4 that the number of protein molecules within a cell decreases only when the cell divides, at which
point that number is roughly halved. With a degradation tag however, protein half-life can be
6 decreased to a range of 40 minutes to ~ 3 hours [104]. Competition for proteases has even been
used as the mechanism for coupling genetic oscillators [105]. While *E. coli* has received the most
8 concerted effort, there have been similar libraries of promoters, RBSs, and degradation tags in
S. cerevisiae yeast [106], [107], [108], *B. subtilis* gram-positive bacteria [109], and mammalian
10 systems [110]. The relative timescales of these essential processes are shown in Figure 7e.

An important step towards predictive models is to ensure that data from individual
12 genetic modules is transferable into larger scale systems. Recently, the developers of Cello
Computer Automated Design (CelloCAD) [111] attempted to encapsulate data from independent
14 characterizations of orthogonal synthetic modules into a predictive model. In [111], the kinetics of
16 NOT and NOR logic gates that depend on a library of 16 orthogonal transcriptional repressors
were characterized. With this data and a small set of modular gates, CelloCAD was used to design
18 a series of multilayered logic gates: 35 out of 52 designed gates worked as predicted the first
time. CelloCAD demonstrated that individual characterization of all components in a library
could be extended to predict multi-layered circuit behavior.

20 In order to escape the costly and slow cycles of transcription and translation, synthetic
circuits would ideally consist of only protein-protein interactions. Phosphorylation-based signal-
22 ing is among the fastest. In [112], synthetic phosphorylation domains were developed with the
purpose of enabling future synthetic circuits to take full advantage of slow and fast time scales.
24 In native cell signaling systems, there will often be fast pathways for information processing
and slow pathways for locking in cell state. The fast pathways for information transduction
26 and processing are mediated by phosphorylation cascades by existing proteins while the slow
pathways will involve gene expression and protein production. Most synthetic circuits to date
28 have relied exclusively on gene expression and protein production, which limits the time scales
at which synthetic circuits can respond. Modular protein domains are engineered such that target
30 proteins only bind to a scaffold once it has been phosphorylated. This system allows for the
design of a number of switches that only respond to certain pulse lengths by taking advantage
32 of fast and slow timescales.

Additional biomolecular parts include quorum sensing systems and DNA recombinases.
34 Quorum sensing systems rely on the diffusion of a small molecule called acyl-homoserine lactone
(AHL). These small molecules diffuse readily across cell membranes, resulting in a response

that is both fast and proportional to the distance from the source. The system requires production of two enzymes: an AHL synthase that makes AHL from a precursor molecule, and an AHL receptor that will initiate gene transcription when activated by AHL. Orthogonal quorum sensing systems can be seen as different “channels” of communication between cells and are critical to the development of synthetic microbial communities [113]. DNA integrases are tetramers that physically invert sequences of DNA – the production of the integrases themselves and then the subsequent DNA flipping contribute to the slow performance tracking. However, permanent DNA recombination can be achieved without constant integrase production, thus the low control effort. This DNA recombination can be used for logic gates[27], [29], finite state machines [26], counters[24], and memory [28]. More in-depth coverage of genetic parts for biomolecular circuits is reviewed in [114].

Controllers in living cells

The state-of-the-art thus far for implementation of biological controllers in synthetic biology can be separated into three categories: synthetic circuits built to demonstrate and characterize specific feedback motifs in a modular fashion, feedback used to regulate metabolic products for increasing titer, and circuits that involve computer-controlled feedback. In this section, we will discuss examples of each as well as the difficulties that arise when implementing these systems in living cells. In many of the examples, we may refer to “heterologous” protein expression or circuits – this refers to the addition of non-native genes to a host organism to enable the production of non-native proteins. The term “toxicity” refers to a spectrum of cellular effects ranging from slowdown in growth, to triggering of stress response pathways, to cell death.

Synthetic circuits for demonstration of feedback modules

A significant portion of studies in synthetic biology has focused on investigating the characteristic properties of feedback modules, as well as a variety of strategies for implementation. These feedback modules include positive and negative feedback, integral control, proportional-integral (PI) control, and feedforward loops (Figure 8a). Commonly, these types of circuits will use fluorescent proteins as outputs, as they can be readily detected during the course of the experiment. Though many of these studies seem simple in their design, they have been instrumental in elucidating biological mechanisms, response times, orthogonality of parts, and unanticipated host interactions.

Negative feedback regulation has been successfully implemented in a variety of ways in synthetic circuits, due to a relative abundance of molecular parts for transcriptional, translational, and protein-protein repression and sequestration. In [115], the use of negative autoregulation with

a strong promoter was shown to significantly improve rise-times towards a desired steady state. [91] used cell-cell communication to implement negative feedback in *E. coli* to regulate the population density. Each bacterium in this system would produce quorum-sensing molecules but also trigger a lysis protein if the internal concentration was too high. In [85], negative feedback on gene expression was implemented using synthetic protein scaffolds to activate or inhibit a two-component system in *E. coli*. The negative feedback loop was then used for real-time fluorescence tracking of a chemical signal in a microfluidics platform. This system had two timescales: the slow transcription/translation time for production of the scaffold, but fast phosphotransfer dynamics.

Positive feedback has been used to rapidly increase response amplitude – usually this is implemented with a transcriptional activator that activates transcription of its own gene. With a highly expressed positive feedback loop, it was demonstrated in [116] that synthetic circuits can drastically disrupt the growth of the host, and that that disruption can feedback onto the performance of the circuit, resulting in bistability in single cell response.

Circuit topologies combining positive and negative feedback have been considered to implement integral and feedforward control. The feedforward loop motif was utilized in [117] to generate pulses in response to an input. Incoherent feedforward loops required both an activator that activates both the output protein and a repressor – as the repressor builds up, the output is throttled until it returns to zero. A quorum sensing system combined with a feedforward loop was used to generate two-dimensional pattern generation in *E. coli* [31]. In [118], both integral and feedforward control were implemented using a quorum-sensing circuit to create scale-invariant patterning. In [119], the characteristics of the basic feedback modules for gene expression in a synthetic system are considered: unregulated gene expression, a repressor-only system, an activator-only system, and a repressor-activator system. By measuring the mean and distribution of *E. coli* cell populations with these modes of gene expression, [119] demonstrated that the response of a combined repressor-activator system with positive feedback could be predicted.

A number of studies have considered many of the theoretical details of integral control [79], [80], [82]. However, thus far synthetic integral control has only been implemented in *in vitro* DNA circuits [120], where reactions occur in a controlled liquid buffer, not in living cells. In [79], an experimentally feasible implementation of PI control with just a repressor and an activator was designed. In addition to simulating perfect adaptation, it was predicted that the system set-point and feedback strength can be tuned by varying the concentration and activity of each protein.

While the synthetic feedback modules discussed in this section have been important

1 for elucidating relative timescales, circuit reliability, and proof-of-concept circuit topologies,
2 reliability and predictability remain persistent challenges. A well-characterized circuit in one
3 strain will behave very differently at a different temperature, in a new host organism, or in
4 fermentation conditions. These difficulties are a result of relatively small numbers of usable,
5 orthogonal parts, challenges with integration of increasingly large numbers of heterologous genes
6 into the cell, and limitations in measurement of more than 3–4 fluorescent outputs. However,
7 new molecular biology tools and libraries of genetic parts are constantly emerging from active
8 research.

Feedback for increased production of metabolic products

10 Metabolic engineering has a singular focus on increasing the titer of a specific molecular
11 product – there are two predominant strategies to achieve this. First, enzymes that produce the
12 product can be up-regulated to increase the overall flux through the cell. Second, titer can be
13 increased by reducing toxicity to the cell caused by the product or its intermediates. Usually,
14 successful engineering of a host for maximum titer involves multiple iterations of both these
15 considerations. Feedback can be utilized to regulate flux through a pathway in a controlled
16 manner that does not allow buildup of toxic or unused intermediate products; it can also be used
17 to export product that does not readily leave the cell, thus preventing additional toxic buildup
18 (Figure 8b).

19 The example of regulating efflux pumps to regulate product buildup has been discussed in a
20 previous section [83]. Optimization of the isoprenoid production pathway is another example. A
21 common metabolic engineering problem is the buildup of toxic intermediate chemicals within the
22 host cell, slowing cell growth and reducing yield of the desired product. In order to dynamically
23 control expression of enzymes along this nine-step pathway, [121] identified transcriptional
24 promoters that would respond proportionally to the concentration of farnesyl pyrophosphate
25 (FPP), one of two toxic intermediates in the pathway. Extensive analysis of whole genome
26 transcripts allowed for identification of a library of native *E. coli* promoters that responded
27 to FPP toxicity; these promoters could then be engineered to drive the entire FPP production
28 pathway, thus enabling dynamic, negative feedback control of FPP levels. Furthermore, it was
29 demonstrated that implementation of these dynamically controlled pathways had much improved
30 performance over systems with open loop (constitutive gene expression) or step induction (small-
31 molecule activated expression) systems of pathway regulation. The flux through an essential
32 metabolic pathway can also be monitored by linking the metabolites to a synthetic oscillator.
33 [122] rewired *E. coli* such that high metabolic flux through the glycolysis pathway would result
34 in fluorescent oscillations that could be observed. This was achieved by linking the promoters for
production and degradation of acetyl-CoA to its product acetyl phosphate. Numerical simulations

and Hopf bifurcation analysis predicted that the system would be stable at low glycolytic flux, and approach a limit cycle as flux increased; Experimentally, the use of different carbon sources with known glycolytic flux levels validated this hypothesis.

Designing reliable switches for metabolic engineering is a particularly challenging problem, since any leaky expression is amplified with the extremely high volumes of commercial production. The goal is to have the production pathways turned off while the cells are dividing, and then, once enough biomass has been generated, to have all the cells turn on. An imperfect OFF-state will result in high mutation rates in the engineered strain or extremely slow growth and insufficient biomass. An imperfect ON-state means that a large number of cells are consuming resources without producing product. Most switch designs in synthetic circuits for demonstration of feedback motifs rely on small-molecule inducers, which are too expensive to be used in large-scale fermentations.

In [123], this issue of metabolic switches was addressed by augmentation of an existing toggle switch design with a glucose-dependent promoter. The classic synthetic toggle switch design uses two repressors that co-repress the reciprocal promoter. Positive feedback results in stochastic bimodality but the switch can be forced one way or the other by inducers that inhibit one of the repressors. In [123], the toggle switch is pre-conditioned in the OFF-state. As the population grows and divides, it uses up all of the available glucose – when the glucose is depleted, this activates a glucose-starvation promoter that flips the toggle switch into the ON-state, and turning on of the production pathway. The toggle retains memory of the ON-state, even when glucose is reintroduced.

The possibility of turning off all RNA polymerase production, and thus all new protein production, was explored in [124] as a strategy for separating the growth phase from the metabolic production phase. [124] replaced the native promoter for the RNA polymerase genes in *E. coli* with an inducible promoter and observed that the growth rate is dependent on the number of polymerases in an ultrasensitive fashion. Above a certain threshold of RNA polymerase, the growth rate of *E. coli* goes from zero to the maximum with a fitted Hill coefficient of 10; Furthermore, wild type levels of polymerase were measured to be just above this threshold, *i.e.* maximizing growth while minimizing the cost of making the polymerase itself. Lastly, it was demonstrated that growth-arrested cells with no new protein production were still able to use their existing enzymes to produce glycerol, suggesting that repression of polymerase upon reaching a certain cell density might be beneficial for fermentation.

Organisms used for large scale production are usually extremely stressed due to the sheer volume of product being produced – this makes the population prime for selection of mutations

that relieve that stress. An ongoing challenge in metabolic engineering is understanding the likely breakage points in a pathway, and also exploring ways of up-regulating production titer without simply increasing the copy number of every enzyme in the pathway. Implementations of feedback loops usually involve adding more genes that need to be expressed in addition to the dozen or so genes that are already being up-regulated for production purposes. The result is that feedback to regulate titer often needs additional safeguards to prevent the feedback from breaking. These are challenges that would benefit from experimentally informed models and a more sophisticated control theoretic approach.

In silico model predictive control of living cells

A number of groups have demonstrated *in silico* feedback systems in living cells. These systems typically involve a custom experimental setup in which the cells are cultured in a platform that allows cell state to be measured in pre-determined intervals - this information is then fed into an algorithm which has real-time control over the cell culture environment (Figure 8c).

For instance, [95] demonstrated a microfluidics platform for *in silico* control of the osmotic stress response in *S. cerevisiae* yeast. This system features a Kalman filter connected to a model predictive control (MPC) system. Yeast have several native fast and slow timescale negative feedback loops that ensure perfect adaptation to osmotic stress. The robustness of the controller was demonstrated by not explicitly modeling these pathways. Instead, a simple two-variable delay differential equation system was used, where osmotic stress was the first variable and fluorescent protein level was the second. Real time image analysis provided the data to calculate the optimal number of osmotic shocks to apply within a 2 hour window. Then, only the first shock is applied and the window is recalculated after six minutes. A toy example of biological model predictive control based on [95] is given in “Sidebar: Model Predictive Control of Gene Expression.” The experimental setup had a yeast monolayer culture in a custom microfluidic device mounted onto an automated microscope for imaging. Osmotic stress response was quantified by stress-induced production of fluorescent proteins inside the yeast nucleus. The brightness of these proteins can be easily imaged and quantified. The fluorescence levels were then fed into the controller and compared to previous time points and set point levels. Osmotic stress levels were modulated by dynamic switching between just two types of media: regular and high salt media.

Similarly, [94] designed an experimental system using light-inducible promoters for the production of a fluorescent protein in *S. cerevisiae* combined with flow cytometry and *in silico* controllers. This setup enabled automated sampling of liquid cultures into a flow cytometer that would measure single-cell fluorescence levels across a population. Similar to [95], the system

used a Kalman filter in conjunction with MPC to estimate unmeasured states in the model, calculate a series of light pulse inputs needed to achieve the desired fluorescence set point, and then apply only the first pulse prior to recalculating at the next time point. By sampling every 30 minutes and applying light pulses every 15 minutes, single-cell level tracking of the desired fluorescence setpoint was demonstrated. In [96], an updated system that could achieve perfect adaptation of *E. coli* to perturbations in nutrient levels and temperature was published. This study also compared experimental results for open loop, PI, and MPC control of not only fluorescent protein levels, but also as applied to cellular growth rates.

MPC of living cells has the advantage of running for long periods of time, and can be used as a precise way of probing endogenous cellular pathways and synthetic circuit performance. However, only the fluorescence output of a few selected nodes in the synthetic circuit can be observed, presenting a major challenge in inferring rates for other parts of the system. Furthermore, a fairly sophisticated experimental setup is needed to successfully run these types of experiments. In particular, smooth transitions between the computer, the microfluidic device, and the image analysis can be notoriously difficult.

Resource limitations and retroactivity

Biomolecular gene circuits are subject to influence from genetic and environmental context. For example, in a phenomenon known as “retroactivity,” loading effects from downstream gene circuit components can affect the performance of upstream components [125], [126]. In addition, genes in a synthetic circuit almost always require access to a common pool of cellular resources, including ribosomes and RNA polymerases, in order to be expressed. The competition for these shared resources can cause unwanted interactions between circuit components via resource competition. Here, we discuss both of these issues along with efforts to alleviate their impact on synthetic circuits.

Retroactivity in genetic circuits is similar to the impedance effects in electrical circuits, where downstream modules can affect the behavior of upstream modules. In electrical circuits, one solution to mitigate these loading effects is to separate the upstream and downstream modules with an operational amplifier that acts as a unity gain buffer. The op amp uses high gain feedback to buffer the upstream module from impedance effects introduced by the downstream module. However, there are often bandwidth limitations on the dynamics of the upstream and downstream modules connected to op amps to ensure that there is sufficient time scale separation between module dynamics and op amp dynamics for that op amp to behave well as a unity gain buffer.

An analogous high gain feedback system based on phosphorylation cycles to attenuate

retroactivity in synthetic gene circuits has been proposed and demonstrated in [127]. Typically, the upstream and downstream modules in synthetic gene circuits are based on gene expression, which occurs on the time scale of minutes. However, phosphorylation cycles occur on the time scale of seconds, which ensures that there is sufficient time scale separation between the load buffering module and the upstream and downstream modules for good buffering performance. This type of system has been implemented in both bacteria [127] and yeast [128]. For a more detailed discussion of phosphorylation based load insulators, see “Sidebar: Attenuation of Loading Effects in Synthetic Gene Circuits.”

A proposed solution to resource competition is to design active resource allocation within the cell, and to have transcriptional machinery that is completely orthogonal to that of the native host. This approach was explored in [129] through implementation of a two-part non-native RNA polymerase in *E. coli* that requires both a main ‘core’ fragment and a smaller ‘targeting’ fragment known as a σ factor. Two orthogonal ‘targeting’ fragments were then engineered, allowing for σ factor specific gene expression. Expansion of this system would allow for controllable allocation of transcriptional resources through regulation of different ‘targeting’ σ factors. Furthermore, additional regulatory fragments for positive and negative control of the shared polymerase were also demonstrated.

As our understanding of biological hosts for synthetic circuits becomes more nuanced, synthetic circuit designs will need to compensate for common resource pools and the coupled effects of retroactivity and loading. This will be essential for increasing the complexity and number of circuits that can be accommodated within a single cell, and requires dynamic regulation of gene expression and resource allocation.

Conclusion: Prevailing Challenges and Open Problems

The first two decades of synthetic biology research have largely focused on the development of genetic engineering tools, biological understanding of host organisms, and the proof-of-concept demonstration of single function synthetic gene circuits. Although the goal of these synthetic gene circuits is to have stackable, modular components, the field has not yet been able to achieve the kind of abstraction needed to increase the complexity of synthetic circuits beyond 1–3 layers of signal transduction.

There are a number of reasons for this – first, there is still a limited toolbox of genetic tools with which we can manipulate DNA, control signal transduction, and maintain memory. Orthogonality of parts is crucial since all genetic components share the same cellular environment and true spatial separation is difficult. This challenge is largely dependent on biological discovery

or cycles of engineering on existing parts. Secondly, although the tools have grown quite
2 advanced, it is still slow to experimentally engineer organisms with large numbers of heterologous
genes. Many efforts are being made to parallelize and speed up large scale engineering of
4 organisms, but it largely remains a low throughput process. This means that the design-build-
test cycle time is on the order of weeks to months, instead of hours to days.

6 Context dependence effects pervade the implementation and performance of synthetic
circuits. The performance specifications for synthetic gene circuits and components change
8 significantly with variations in parameters such as temperature, host organism, growth media
formulation, and position of the genes in the genome. Feedback control, which generally
10 improves system robustness, has the potential to improve the robustness of synthetic circuits
across these varying contexts. To conclude this work, we briefly review some of the open
12 problems in system identification, controller design, and control implementation in synthetic
biology.

14 Because of context dependence, system identification is especially crucial for synthetic
gene circuits. Unlike mechanical or electrical systems, which can be modeled universally to
16 some degree using physics, synthetic gene circuits must be characterized in their desired context.
Models for synthetic gene circuits often contain very many parameters and species to describe
18 a low dimensional output. These large models can suffer from parameter degeneracy as well
as computational cost. The approaches to alleviate these issues could include model reduction
20 or computational speed-ups [71], [72]. More development in these areas will help make system
identification more tractable for more complex synthetic circuits.

22 In control design, the challenges include stochastic distributional control of populations
and the development of an algorithmic optimal control procedure. Populations often exhibit bet-
24 hedging behaviors such as bacterial persistence (see Figure 3b) in which some small percentage
of cells prepare for a catastrophic event that may or may not occur. In this way, some cells will
26 be able to survive such an event. In this type of behavior, the steady state of the population is
not a set of species concentrations inside one cell, but rather it is a steady state distribution of
28 species concentrations across the population. An open area in control theory is the development
of design tools for creating synthetic circuits that create a robust distribution of an output across
30 a population.

Another major open challenge is the development of optimal control for synthetic gene
32 circuits in a manner analogous to H_2 or H_∞ optimal control. In [83], the authors compare
different control architectures for implementation of negative feedback by creating different
34 models and comparing performance using simulations. Additionally, as shown in Figure 7,

feedback can be implemented at different layers in the cell. An optimal control algorithm
2 that could take some control cost function and automatically produce a control architecture
minimizing that cost function using feedback at the appropriate layers would help accelerate the
4 design process.

A third open challenge is to develop unifying principles for design of feedback gene
6 circuits that mitigate context dependence effects. There are many feedback systems that mitigate
context dependence of a certain type, but ideally, synthetic gene circuits would be robust to
8 multiple difference context effects. For example, a gene circuit should work while varying both
growth medium and temperature simultaneously. While this may seem as simple as combining
10 a feedback loop for temperature robustness together with one for growth medium robustness,
scaling such an approach up will require careful consideration of both resource limits and the
12 architecture with which to combine multiple feedback loops.

Design of feedback with resource limits in mind could be especially useful in metabolic
14 engineering, where feedback can be used to improve yield. However, adding a controller also
forces the cell to spend energy and material making the controller species, which could divert
16 flux away from the desired metabolic pathway and reduce performance. Therefore, designing
efficient genetic controllers could be important.

18 Finally, the challenges in implementation of controllers greatly overlap with the general
challenges facing synthetic biology. That is, implementation of complex synthetic circuits is
20 difficult because of limited characterized and orthogonal parts, slow design build test cycles, and
context dependence effects. Specifically, control performance is almost always limited by sensing
22 and actuation capability. Thus, developing more synthetic genetic parts for sensing different
outputs as well as faster actuators that can reject higher frequency disturbances will help push
24 the limits of feedback control in single cells.

As scientists slowly peel back the complex layers of regulation that govern the everyday
26 survival and function of living cells, it has become abundantly clear that interdisciplinary research
is necessary in order to integrate all of the data from individual protein function to systems level
28 networks, and to truly understand the cell as a dynamic system. Synthetic biology presents unique
challenges for modeling and control; the development of biology-informed control principles is
2 necessary to take the field to the next level of complexity.

References

- 4 [1] D. E. Cameron, C. J. Bashor, and J. J. Collins, “A brief history of synthetic biology.” *Nature Reviews Microbiology*, vol. 12, no. 5, pp. 381–390, May 2014.
- 6 [2] S. K. Lee, H. Chou, T. S. Ham, T. S. Lee, and J. D. Keasling, “Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels.” *Current Opinion in Biotechnology*, vol. 19, no. 6, pp. 556–563, Dec. 2008.
- 8 [3] J. D. Keasling, “Synthetic biology and the development of tools for metabolic engineering.” *Metabolic Engineering*, vol. 14, no. 3, pp. 189–195, May 2012.
- 10 [4] T. K. Lu, A. S. Khalil, and J. J. Collins, “Next-generation synthetic gene networks.” *Nature Biotechnology*, vol. 27, no. 12, pp. 1139–1150, Dec. 2009.
- 12 [5] A. S. Khalil and J. J. Collins, “Synthetic biology: applications come of age.” *Nature Reviews Genetics*, vol. 11, no. 5, pp. 367–379, May 2010.
- 14 [6] W. C. Ruder, T. Lu, and J. J. Collins, “Synthetic biology moving into the clinic.” *Science*, vol. 333, no. 6047, pp. 1248–1252, Sep. 2011.
- 16 [7] K. Brenner, L. You, and F. H. Arnold, “Engineering microbial consortia: a new frontier in synthetic biology.” *Trends in Biotechnology*, vol. 26, no. 9, pp. 483–489, Sep. 2008.
- 18 [8] J. Shong, M. R. Jimenez Diaz, and C. H. Collins, “Towards synthetic microbial consortia for bioprocessing.” *Current Opinion in Biotechnology*, vol. 23, no. 5, pp. 798–802, Oct. 2012.
- 20 [9] D. Del Vecchio, A. J. Dy, and Y. Qian, “Control theory meets synthetic biology.” *Journal of The Royal Society Interface*, vol. 13, no. 120, p. 20160380, Jul. 2016.
- 22 [10] T. P. Prescott and A. Papachristodoulou, “Synthetic biology: A control engineering perspective,” in *2014 European Control Conference (ECC)*. IEEE, 2014, pp. 1182–1186.
- 24 [11] M. E. Csete and J. C. Doyle, “Reverse engineering of biological complexity.” *Science*, vol. 295, no. 5560, pp. 1664–1669, Mar. 2002.
- 26 [12] M. B. Elowitz and S. Leibler, “A synthetic oscillatory network of transcriptional regulators.” *Nature*, vol. 403, no. 6767, pp. 335–338, Jan. 2000.
- 28 [13] T. S. Gardner, C. R. Cantor, and J. J. Collins, “Construction of a genetic toggle switch in *Escherichia coli*.” *Nature*, vol. 403, no. 6767, pp. 339–342, Jan. 2000.
- 30 [14] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, “A fast, robust and tunable synthetic gene oscillator,” *Nature*, vol. 456, no. 7221, pp. 516–519, Oct. 2008.
- 32 [15] A. Prindle, P. Samayoa, I. Razinkov, T. Danino, L. S. Tsimring, and J. Hasty, “A sensing array of radically coupled genetic/biopixels/,” *Nature*, vol. 481, no. 7379, pp. 39–44, 2012.
- 34 [16] H. Niederholtmeyer, Z. Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, S. J.

- Maerkl, and F. Simmel, “Rapid cell-free forward engineering of novel genetic ring oscillators,” *eLife*, vol. 4, p. e09771, Dec. 2015.
- [17] L. Potvin-Trottier, N. D. Lord, G. Vinnicombe, and J. Paulsson, “Synchronous long-term oscillations in a synthetic gene circuit.” *Nature*, vol. 538, no. 7626, pp. 514–517, Oct. 2016.
- [18] A. A. Green, P. A. Silver, J. J. Collins, and P. Yin, “Toehold Switches: De-Novo-Designed Regulators of Gene Expression,” *Cell*, vol. 159, no. 4, pp. 925–939, Nov. 2014.
- [19] J. C. Anderson, C. A. Voigt, and A. P. Arkin, “Environmental signal integration by a modular AND gate.” *Molecular Systems Biology*, vol. 3, p. 133, 2007.
- [20] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt, “Genetic programs constructed from layered logic gates in single cells,” *Nature*, vol. 491, no. 7423, pp. 249–253, Apr. 2013.
- [21] D. L. Shis, F. Hussain, S. Meinhardt, L. Swint-Kruse, and M. R. Bennett, “Modular, Multi-Input Transcriptional Logic Gating with Orthogonal LacI/GalR Family Chimeras,” *ACS Synthetic Biology*, p. 140728104557004, Jul. 2014.
- [22] B. Wang, R. I. Kitney, N. Joly, and M. Buck, “Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology,” *Nature Communications*, vol. 2, p. 508, Oct. 2011.
- [23] V. Hsiao, Y. Hori, P. W. Rothmund, and R. M. Murray, “A population-based temporal logic gate for timing and recording chemical events.” *Molecular Systems Biology*, vol. 12, no. 5, p. 869, 2016.
- [24] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins, “Synthetic gene networks that count.” *Science*, vol. 324, no. 5931, pp. 1199–1202, May 2009.
- [25] K. Oishi and E. Klavins, “Framework for engineering finite state machines in gene regulatory networks.” *ACS Synthetic Biology*, vol. 3, no. 9, pp. 652–665, Sep. 2014.
- [26] N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, and T. K. Lu, “Synthetic recombinase-based state machines in living cells,” *Science*, vol. 353, no. 6297, p. aad8559, Jul. 2016.
- [27] P. Siuti, J. Yazbek, and T. K. Lu, “Synthetic circuits integrating logic and memory in living cells.” *Nature Biotechnology*, vol. 31, no. 5, pp. 448–452, May 2013.
- [28] L. Yang, A. A. K. Nielsen, J. Fernandez-Rodriguez, C. J. McClune, M. T. Laub, T. K. Lu, and C. A. Voigt, “Permanent genetic memory with >1-byte capacity.” *Nature Methods*, vol. 11, no. 12, pp. 1261–1266, Dec. 2014.
- [29] J. Bonnet, P. Subsoontorn, and D. Endy, “Rewritable digital data storage in live cells via engineered control of recombination directionality.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 23, pp. 8884–8889, Jun. 2012.
- [30] F. K. Balagaddé, H. Song, J. Ozaki, C. H. Collins, M. Barnet, F. H. Arnold, S. R. Quake,

- and L. You, “A synthetic *Escherichia coli* predator-prey ecosystem.” *Molecular Systems Biology*, vol. 4, p. 187, 2008.
- [31] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, “A synthetic multicellular system for programmed pattern formation.” *Nature*, vol. 434, no. 7037, pp. 1130–1134, Apr. 2005.
- [32] C. J. Paddon and J. D. Keasling, “Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development,” *Nature Reviews Microbiology*, vol. 12, no. 5, pp. 355–367, Apr. 2014.
- [33] S. S. Chandran, J. T. Kealey, and C. D. Reeves, “Microbial production of isoprenoids,” *Process Biochemistry*, vol. 46, no. 9, pp. 1703–1710, 2011.
- [34] A. Burgard, M. J. Burk, R. Osterhout, S. Van Dien, and H. Yim, “Development of a commercial scale process for production of 1,4-butanediol from sugar.” *Current Opinion in Biotechnology*, vol. 42, pp. 118–125, Dec. 2016.
- [35] T. Scheibel, “Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins.” *Microbial cell factories*, vol. 3, no. 1, p. 14, Nov. 2004.
- [36] F. Ceroni, R. Algar, G.-B. Stan, and T. Ellis, “Quantifying cellular capacity identifies gene expression designs with reduced burden.” *Nature Publishing Group*, vol. 12, no. 5, pp. 415–418, May 2015.
- [37] S. C. Sleight and H. M. Sauro, “Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits.” *ACS Synthetic Biology*, vol. 2, no. 9, pp. 519–528, Sep. 2013.
- [38] O. Wright, M. Delmans, G.-B. Stan, and T. Ellis, “GeneGuard: A modular plasmid system designed for biosafety.” *ACS Synthetic Biology*, vol. 4, no. 3, pp. 307–316, Mar. 2015.
- [39] O. Wright, G.-B. Stan, and T. Ellis, “Building-in biosafety for synthetic biology.” *Microbiology*, vol. 159, no. Pt 7, pp. 1221–1235, Jul. 2013.
- [40] H. El-Samad, J. P. Goff, and M. Khammash, “Calcium homeostasis and parturient hypocalcemia: an integral feedback perspective.” *Journal of Theoretical Biology*, vol. 214, no. 1, pp. 17–29, Jan. 2002.
- [41] T.-M. Yi, Y. Huang, M. I. Simon, and J. Doyle, “Robust perfect adaptation in bacterial chemotaxis through integral feedback control,” *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4649–4653, 2000.
- [42] W. Ma, A. Trusina, H. El Samad, W. A. Lim, and C. Tang, “Defining network topologies that can achieve biochemical adaptation.” *Cell*, vol. 138, no. 4, pp. 760–773, Aug. 2009.
- [43] D. Muzzey, C. A. Gmez-Urbe, J. T. Mettetal, and A. van Oudenaarden, “A systems-level analysis of perfect adaptation in yeast osmoregulation,” *Cell*, vol. 138, no. 1, pp. 160–171, 2009.

- 4 [44] J. E. Ferrell, “Perfect and Near-Perfect Adaptation in Cell Signaling.” *Cell Systems*, vol. 2, no. 2, pp. 62–67, Feb. 2016.
- 6 [45] J. L. Radzikowski, S. Vedelaar, D. Siegel, Á. D. Ortega, A. Schmidt, and M. Heinemann, “Bacterial persistence is an active σ S stress response to metabolic flux limitation.” *Molecular Systems Biology*, vol. 12, no. 9, p. 882, Sep. 2016.
- 8 [46] L. Gelens, G. A. Anderson, and J. E. Ferrell, “Spatial trigger waves: positive feedback gets you a long way.” *Molecular Biology of the Cell*, vol. 25, no. 22, pp. 3486–3493, Nov. 2014.
- 10 [47] O. S. Venturelli, H. El-Samad, and R. M. Murray, “Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 48, pp. E3324–E3333, 2012.
- 12 [48] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, “An excitable gene regulatory circuit induces transient cellular differentiation.” *Nature*, vol. 440, no. 7083, pp. 545–550, Mar. 2006.
- 14 [49] S. Zeng, S. Waldherr, and F. Allgöwer, “An inverse problem of tomographic type in population dynamics,” in *53rd IEEE Conference on Decision and Control*, 2014, pp. 1643–1648.
- 16 [50] A. Swaminathan and R. M. Murray, “Linear system identifiability from distributional and time series data,” in *American Control Conference (ACC)*, 2016, pp. 392–399.
- 18 [51] B. Munsky and M. Khammash, “Identification from stochastic cell-to-cell variation: a genetic switch case study,” *IET Systems Biology*, vol. 4, no. 6, pp. 356–366(10), 2010.
- 20 [52] G. Lillacci and M. Khammash, “Parameter estimation and model selection in computational biology,” *PLOS Computational Biology*, vol. 6, no. 3, pp. 1–17, 2010.
- 22 [53] Y. Hori and R. M. Murray, “A state-space realization approach to set identification of biochemical kinetic parameters,” in *2015 European Control Conference (ECC)*, 2015, pp. 2280–2285.
- 24 [54] J. Hasenauer, S. Waldherr, K. Wagner, and F. Allgöwer, “Parameter identification, experimental design and model falsification for biological network models using semidefinite programming,” *IET Systems Biology*, vol. 4, pp. 119–130(11), 2010.
- 26 [55] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf, “Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems,” *Journal of The Royal Society Interface*, vol. 6, no. 31, pp. 187–202, 2009.
- 28 [56] W. Pan, Y. Yuan, J. Goncalves, and G.-B. Stan, “A Sparse Bayesian Approach to the Identification of Nonlinear State-Space Systems,” *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 182–187, 2016.
- 30 [57] B. Munsky and M. Khammash, “The finite state projection algorithm for the solution
- 32
- 34
- 36
- 2

of the chemical master equation,” *The Journal of Chemical Physics*, vol. 124, no. 4, p. 044104, 2006.

- [58] G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, “Systematic identification of signal-activated stochastic gene regulation,” *Science*, vol. 339, no. 6119, pp. 584–587, 2013.
- [59] B. Munsky, B. Trinh, and M. Khammash, “Listening to the noise: random fluctuations reveal gene network parameters,” *Molecular Systems Biology*, vol. 5, no. 1, pp. 1–7, 2009.
- [60] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koepl, “Moment-based inference predicts bimodality in transient gene expression,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 21, pp. 8340–8345, 2012.
- [61] J. Lipinski-Kruszka, J. Stewart-Ornstein, M. W. Chevalier, and H. El-Samad, “Using dynamic noise propagation to infer causal regulatory relationships in biochemical networks,” *ACS Synthetic Biology*, vol. 4, no. 3, pp. 258–264, 2015.
- [62] J. Ruess, F. Parise, A. Miliadis-Argeitis, M. Khammash, and J. Lygeros, “Iterative experiment design guides the characterization of a light-inducible gene expression circuit.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 26, pp. 8148–8153, Jun. 2015.
- [63] A. Singh and J. P. Hespanha, “Approximate moment dynamics for chemically reacting systems,” *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 414–418, 2011.
- [64] M. W. Chevalier and H. El-Samad, “A data-integrated method for analyzing stochastic biochemical networks,” *The Journal of Chemical Physics*, vol. 135, no. 21, p. 214110, 2011.
- [65] J. Kuntz, M. Ottobre, G.-B. Stan, and M. Barahona, “Bounding Stationary Averages of Polynomial Diffusions via Semidefinite Programming,” *SIAM Journal on Scientific Computing*, vol. 38, no. 6, pp. A3891–A3920, 2016.
- [66] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, “Population growth of human y chromosomes: a study of y chromosome microsatellites.” *Molecular Biology and Evolution*, vol. 16, no. 12, p. 1791, 1999.
- [67] G. Lillacci and M. Khammash, “The signal within the noise: efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations,” *Bioinformatics*, vol. 29, no. 18, p. 2311, 2013.
- [68] J. Hasenauer, S. Waldherr, M. Doszczak, N. Radde, P. Scheurich, and F. Allgöwer, “Identification of models of heterogeneous cell populations from population snapshot data,” *BMC Bioinformatics*, vol. 12, no. 1, p. 125, 2011.
- [69] J. Hasenauer, C. Hasenauer, T. Hucho, and F. J. Theis, “Ode constrained mixture modelling: A method for unraveling subpopulation structures and dynamics,” *PLOS Computational Biology*, vol. 10, no. 7, pp. 1–17, 2014.

- [70] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavar, “Markov chain monte carlo without likelihoods,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 26, pp. 15 324–15 328, 2003.
- [71] A. Golightly and D. J. Wilkinson, “Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo,” *Interface Focus*, vol. 1, no. 6, pp. 807–820, 2011.
- [72] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koepl, “Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings,” *Nature Methods*, vol. 11, no. 2, pp. 197–202, 2014.
- [73] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback control theory*. Courier Corporation, 2013.
- [74] A. Mitchell, P. Wei, and W. A. Lim, “Oscillatory stress stimulation uncovers an achilles’ heel of the yeast mapk signaling network,” *Science*, vol. 350, no. 6266, pp. 1379–1383, 2015.
- [75] F. A. Chandra, G. Buzi, and J. C. Doyle, “Glycolytic oscillations and limits on robust efficiency,” *Science*, vol. 333, no. 6039, pp. 187–192, 2011.
- [76] I. Lestas, G. Vinnicombe, and J. Paulsson, “Fundamental limits on the suppression of molecular fluctuations,” *Nature*, vol. 467, no. 7312, pp. 174–178, 09 2010.
- [77] V. K. Mutalik, J. C. Guimaraes, G. Cambray, C. Lam, M. J. Christoffersen, Q.-A. Mai, A. B. Tran, M. Paull, J. D. Keasling, A. P. Arkin, and D. Endy, “Precise and reliable gene expression via standard transcription and translation initiation elements,” *Nature Methods*, vol. 10, no. 4, pp. 354–360, 2013.
- [78] J. A. J. Arpino, E. J. Hancock, J. Anderson, M. Barahona, G.-B. V. Stan, A. Papachristodoulou, and K. Polizzi, “Tuning the dials of Synthetic Biology.” *Microbiology*, vol. 159, no. Pt 7, pp. 1236–1253, Jul. 2013.
- [79] J. Ang, S. Bagh, B. P. Ingalls, and D. R. McMillen, “Considerations for using integral feedback control to construct a perfectly adapting synthetic gene network.” *Journal of Theoretical Biology*, vol. 266, no. 4, pp. 723–738, Oct. 2010.
- [80] J. Ang and D. R. McMillen, “Physical constraints on biological integral control design for homeostasis and sensory adaptation.” *Biophysical journal*, vol. 104, no. 2, pp. 505–515, Jan. 2013.
- [81] X. Y. Ni, T. Drengstig, and P. Ruoff, “The control of the controller: Molecular mechanisms for robust perfect adaptation and temperature compensation,” *Biophysical Journal*, vol. 97, no. 5, pp. 1244–1253, 2009.
- [82] C. Briat, A. Gupta, and M. Khammash, “Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks.” *Cell Systems*, vol. 2, no. 1, pp. 15–26, Jan. 2016.

- 4 [83] M. J. Dunlop, J. D. Keasling, and A. Mukhopadhyay, “A model for improving microbial
biofuel production using a synthetic feedback loop,” *Systems and Synthetic Biology*, vol. 4,
no. 2, pp. 95–104, 2010.
- 6 [84] Y. Qian, H.-H. Huang, J. I. Jimnez, and D. Del Vecchio, “Resource competition shapes
the response of genetic circuits,” *ACS Synthetic Biology*, vol. 6, no. 7, pp. 1263–1272,
8 2017.
- [85] V. Hsiao, E. L. C. de los Santos, W. R. Whitaker, J. E. Dueber, and R. M. Murray,
10 “Design and implementation of a biomolecular concentration tracker.” *ACS Synthetic
Biology*, vol. 4, no. 2, pp. 150–161, Feb. 2015.
- 12 [86] D. A. Oyarzún and G.-B. V. Stan, “Synthetic gene circuits for metabolic control: design
trade-offs and constraints,” *Journal of The Royal Society Interface*, vol. 10, no. 78, 2012.
- 14 [87] D. A. Oyarzún, J.-B. Lugagne, and G.-B. V. Stan, “Noise propagation in synthetic gene
circuits for metabolic control.” *ACS Synthetic Biology*, vol. 4, no. 2, pp. 116–125, Feb.
16 2015.
- [88] E. V. Nikolaev and E. D. Sontag, “Quorum-sensing synchronization of synthetic toggle
18 switches: A design based on monotone dynamical systems theory,” *PLOS Computational
Biology*, vol. 12, no. 4, pp. 1–33, 2016.
- 20 [89] M. W. Hirsch and H. Smith, “Monotone dynamical systems,” *Handbook of differential
equations: ordinary differential equations*, vol. 2, pp. 239–357, 2006.
- 22 [90] M. Arcak, “Pattern formation by lateral inhibition in large-scale networks of cells,” *IEEE
Transactions on Automatic Control*, vol. 58, no. 5, pp. 1250–1262, 2013.
- 24 [91] L. You, R. S. Cox, R. Weiss, and F. H. Arnold, “Programmed population control by cell-
cell communication and regulated killing.” *Nature*, vol. 428, no. 6985, pp. 868–871, Apr.
26 2004.
- [92] X. Ren, A.-A. Baetica, A. Swaminathan, and R. M. Murray, “Population regulation in
28 microbial consortia using dual feedback control,” *bioRxiv*, 2017.
- [93] C. E. Garca, D. M. Prett, and M. Morari, “Model predictive control: Theory and practice
30 survey,” *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.
- [94] A. Miliás-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El Samad,
32 M. Khammash, and J. Lygeros, “In silico feedback for in vivo regulation of a gene
expression circuit.” *Nature Biotechnology*, vol. 29, no. 12, pp. 1114–1116, Dec. 2011.
- 34 [95] J. Uhlenhof, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and
P. Hersen, “Long-term model predictive control of gene expression at the population and
36 single-cell levels.” *Proceedings of the National Academy of Sciences of the United States
of America*, vol. 109, no. 35, pp. 14 271–14 276, Aug. 2012.
- 2 [96] A. Miliás-Argeitis, M. Rullan, S. K. Aoki, P. Buchmann, and M. Khammash, “Automated

- optogenetic feedback control for precise and robust regulation of gene expression and cell growth.” *Nature Communications*, vol. 7, p. 12546, 2016.
- [97] F. Menolascina, G. Fiore, E. Orabona, L. De Stefano, M. Ferry, J. Hasty, M. di Bernardo, and D. di Bernardo, “In-vivo real-time control of protein expression from endogenous and synthetic gene networks,” *PLOS Computational Biology*, vol. 10, no. 5, pp. 1–14, 2014.
- [98] S. Kiani, A. Chavez, M. Tuttle, R. N. Hall, R. Chari, D. Ter-Ovanesyan, J. Qian, B. W. Pruitt, J. Beal, S. Vora, J. Buchthal, E. J. K. Kowal, M. R. Ebrahimkhani, J. J. Collins, R. Weiss, and G. Church, “Cas9 gRNA engineering for genome editing, activation and repression.” *Nature Publishing Group*, vol. 12, no. 11, pp. 1051–1054, Nov. 2015.
- [99] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, “BioNumbers—the database of key numbers in molecular and cell biology,” *Nucleic Acids Research*, vol. 38, no. Database, pp. D750–D753, 2009.
- [100] R. S. Cox, M. G. Surette, and M. B. Elowitz, “Programming gene expression with combinatorial promoters.” *Molecular Systems Biology*, vol. 3, p. 145, 2007.
- [101] B. C. Stanton, A. A. K. Nielsen, A. Tamsir, K. Clancy, T. Peterson, and C. A. Voigt, “Genomic mining of prokaryotic repressors for orthogonal logic gates,” *Nature Chemical Biology*, vol. 10, no. 2, pp. 99–105, 2014.
- [102] H. M. Salis, E. A. Mirsky, and C. A. Voigt, “Automated design of synthetic ribosome binding sites to control protein expression,” *Nature Biotechnology*, vol. 27, no. 10, pp. 946–950, 2009.
- [103] V. K. Mutalik, J. C. Guimaraes, G. Cambray, Q.-A. Mai, M. J. Christoffersen, L. Martin, A. Yu, C. Lam, C. Rodriguez, G. Bennett, J. D. Keasling, D. Endy, and A. P. Arkin, “Quantitative estimation of activity and quality for collections of functional genetic elements,” *Nature Methods*, vol. 10, no. 4, pp. 347–353, 2013.
- [104] J. B. Andersen, C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin, “New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria.” *Applied and Environmental Microbiology*, vol. 64, no. 6, pp. 2240–2246, Jun. 1998.
- [105] A. Prindle, J. Selimkhanov, H. Li, I. Razinkov, L. S. Tsimring, and J. Hasty, “Rapid and tunable post-translational coupling of genetic circuits.” *Nature*, vol. 508, no. 7496, pp. 387–391, Apr. 2014.
- [106] H. Redden, N. Morse, and H. S. Alper, “The synthetic biology toolbox for tuning gene expression in yeast.” *FEMS Yeast Research*, vol. 15, no. 1, pp. 1–10, Feb. 2015.
- [107] R. M. C. Portela, T. Vogl, C. Kniely, J. E. Fischer, R. Oliveira, and A. Glieder, “Synthetic Core Promoters as Universal Parts for Fine-Tuning Expression in Different Yeast Species.” *ACS Synthetic Biology*, vol. 6, no. 3, pp. 471–484, Mar. 2017.
- [108] A. S. Rajkumar, G. Liu, D. Bergenholm, D. Arsovska, M. Kristensen, J. Nielsen, M. K.

- Jensen, and J. D. Keasling, “Engineering of synthetic, stress-responsive yeast promoters.” *Nucleic Acids Research*, vol. 44, no. 17, pp. e136–e136, Sep. 2016.
- [109] S. Guiziou, V. Sauveplane, H.-J. Chang, C. Clerté, N. Declerck, M. Jules, and J. Bonnet, “A part toolbox to tune genetic expression in *Bacillus subtilis*.” *Nucleic Acids Research*, vol. 44, no. 15, pp. 7495–7508, Sep. 2016.
- [110] A. S. Khalil, T. K. Lu, C. J. Bashor, C. L. Ramirez, N. C. Pyenson, J. K. Joung, and J. J. Collins, “A synthetic biology framework for programming eukaryotic transcription functions.” *Cell*, vol. 150, no. 3, pp. 647–658, Aug. 2012.
- [111] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt, “Genetic circuit design automation.” *Science*, vol. 352, no. 6281, pp. aac7341–aac7341, Apr. 2016.
- [112] R. M. Gordley, R. E. Williams, C. J. Bashor, J. E. Toettcher, S. Yan, and W. A. Lim, “Engineering dynamical control of cell fate switching using synthetic phospho-regulons.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 47, pp. 13 528–13 533, Nov. 2016.
- [113] S. R. Scott and J. Hasty, “Quorum Sensing Communication Modules for Microbial Consortia.” *ACS Synthetic Biology*, vol. 5, no. 9, pp. 969–977, Sep. 2016.
- [114] A. A. K. Nielsen, T. H. Segall-Shapiro, and C. A. Voigt, “Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression.” *Current opinion in chemical biology*, vol. 17, no. 6, pp. 878–892, Dec. 2013.
- [115] N. Rosenfeld, M. B. Elowitz, and U. Alon, “Negative autoregulation speeds the response times of transcription networks.” *Journal of Molecular Biology*, vol. 323, no. 5, pp. 785–793, Nov. 2002.
- [116] C. Tan, P. Marguet, and L. You, “Emergent bistability by a growth-modulating positive feedback circuit,” *Nature Methods*, vol. 5, no. 11, pp. 842–848, Oct. 2009.
- [117] S. Mangan and U. Alon, “Structure and function of the feed-forward loop network motif.” *PNAS*, vol. 100, no. 21, pp. 11 980–11 985, Oct. 2003.
- [118] Y. Cao, M. D. Ryser, S. Payne, B. Li, C. V. Rao, and L. You, “Collective Space-Sensing Coordinates Pattern Scaling in Engineered Bacteria.” *Cell*, vol. 165, no. 3, pp. 620–630, Apr. 2016.
- [119] N. J. Guido, X. Wang, D. Adalsteinsson, D. McMillen, J. Hasty, C. R. Cantor, T. C. Elston, and J. J. Collins, “A bottom-up approach to gene regulation.” *Nature*, vol. 439, no. 7078, pp. 856–860, Feb. 2006.
- [120] C. Briat, C. Zechner, and M. Khammash, “Design of a Synthetic Integral Feedback Circuit: Dynamic Analysis and DNA Implementation,” *ACS Synthetic Biology*, 2016.
- [121] R. H. Dahl, F. Zhang, J. Alonso-Gutierrez, E. Baidoo, T. S. Batth, A. M. Redding-Johanson, C. J. Petzold, A. Mukhopadhyay, T. S. Lee, P. D. Adams, and J. D.

- Keasling, “Engineering dynamic pathway regulation using stress-response promoters.” *Nature Biotechnology*, vol. 31, no. 11, pp. 1039–1046, Nov. 2013.
- [122] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S.-g. Lee, and J. C. Liao, “A synthetic gene-metabolic oscillator.” *Nature*, vol. 435, no. 7038, pp. 118–122, May 2005.
- [123] W. Bothfeld, G. Kapov, and K. E. J. Tyo, “A Glucose-Sensing Toggle Switch for Autonomous, High Productivity Genetic Control.” *ACS Synthetic Biology*, Mar. 2017.
- [124] J. Izard, C. D. C. Gomez Balderas, D. Ropers, S. Lacour, X. Song, Y. Yang, A. B. Lindner, J. Geiselman, and H. de Jong, “A synthetic growth switch based on controlled expression of RNA polymerase.” *Molecular Systems Biology*, vol. 11, no. 11, p. 840, Nov. 2015.
- [125] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag, “Modular cell biology: retroactivity and insulation.” *Molecular Systems Biology*, vol. 4, p. 161, 2008.
- [126] S. Jayanthi, K. S. Nilgiriwala, and D. Del Vecchio, “Retroactivity controls the temporal dynamics of gene transcription.” *ACS Synthetic Biology*, vol. 2, no. 8, pp. 431–441, Aug. 2013.
- [127] K. S. Nilgiriwala, J. Jimnez, P. M. Rivera, and D. Del Vecchio, “Synthetic tunable amplifying buffer circuit in *e. coli*,” *ACS Synthetic Biology*, vol. 4, no. 5, pp. 577–584, 2015.
- [128] D. Mishra, P. M. Rivera, A. Lin, D. Del Vecchio, and R. Weiss, “A load driver device for engineering modularity in biological networks.” *Nature Biotechnology*, vol. 32, no. 12, pp. 1268–1275, Dec. 2014.
- [129] T. H. Segall-Shapiro, A. J. Meyer, A. D. Ellington, E. D. Sontag, and C. A. Voigt, “A ‘resource allocator’ for transcription based on a highly fragmented T7 RNA polymerase.” *Molecular Systems Biology*, vol. 10, no. 7, pp. 742–742, Jul. 2014.
- [130] S. Cardinale and A. P. Arkin, “Contextualizing context for synthetic biology identifying causes of failure of synthetic biological systems,” *Biotechnology Journal*, vol. 7, no. 7, pp. 856–866, 2012.
- [131] D. T. Gillespie, “Exact stochastic simulation of coupled chemical reactions,” *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, Dec. 1977.
- [132] N. G. van Kampen, “Stochastic processes in physics and chemistry,” 1995.
- [133] D. A. McQuarrie, “Stochastic approach to chemical kinetics,” *Journal of Applied Probability*, vol. 4, no. 3, pp. 413–478, 1967.
- [134] A. Swaminathan, V. Hsiao, and R. M. Murray, “Quantitative modeling of integrase dynamics using a novel Python toolbox for parameter inference in synthetic biology,” *bioRxiv*, 2017.
- [135] MATLAB, *version 9.0.0 (R2016a)*. Natick, Massachusetts: The MathWorks Inc., 2016.
- [136] T. R. Maarleveld, B. G. Olivier, and F. J. Bruggeman, “Stochpy: A comprehensive, user-

friendly tool for simulating stochastic biological processes,” *PLOS ONE*, vol. 8, no. 11,
2 pp. 1–10, 2013.

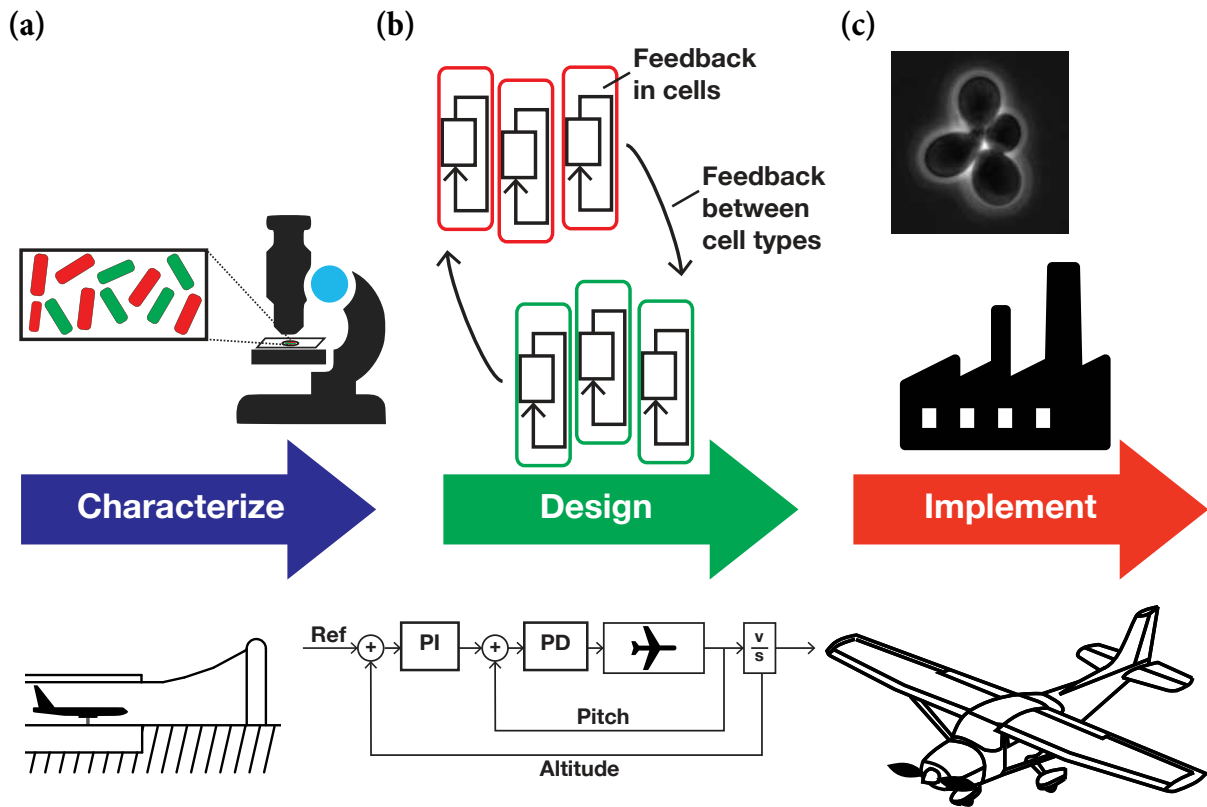


Figure 1. The traditional workflow in control engineering (bottom row) contrasted with engineering of biological systems (top row). (a) The system must be characterized and identified. Microbial systems can be characterized by assays such as microscopy, while a car or plane might be tested in a wind tunnel. (b) A controller must be designed. Having multiple control loops operating at different spatiotemporal scales is common in man-made systems. In engineered biological systems, feedback with nested loops is not as tractable. (c) The controller is then implemented. While modern aviation technology is largely autonomous, an engineered yeast strain must be grown in precisely controlled fermentation tanks to produce the desired molecular products.

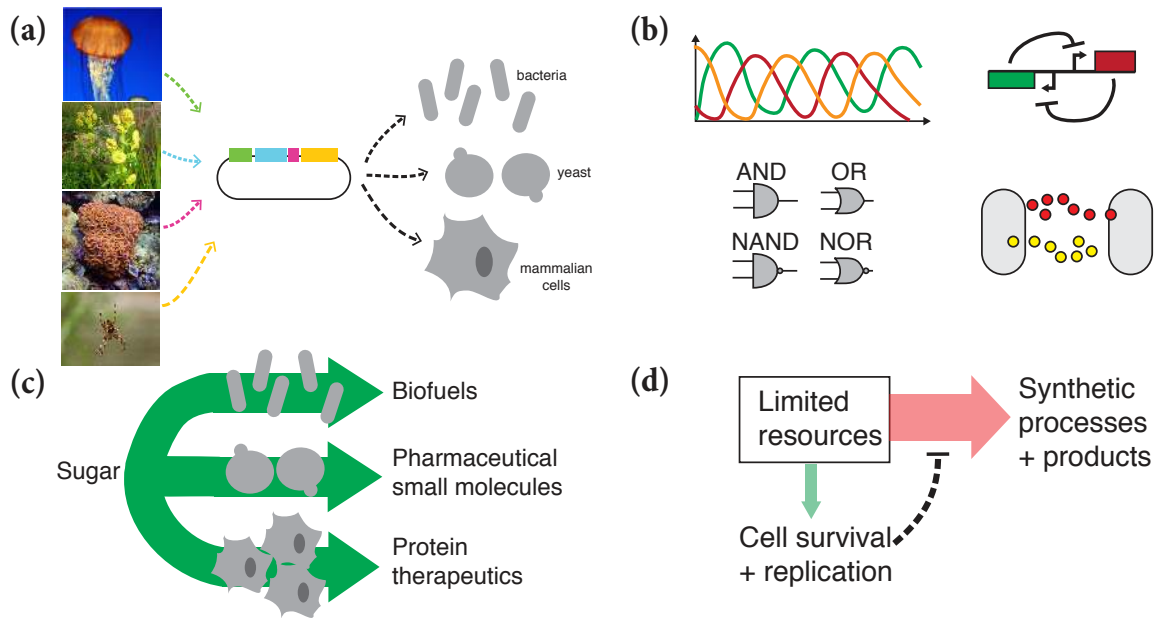


Figure 2. Synthetic biology repurposes genes for new functions. (a) Modern molecular biology enables transplanting genes from many different organisms into new host organisms. Example host chassis organisms that are used most commonly, and thus are the best understood, are *E. coli* or *B. subtilis* bacteria, *S. cerevisiae* yeast, and Chinese Hamster Ovary (CHO) mammalian cells. (b) Synthetic gene circuits have been built to demonstrate oscillators, toggle switches, logic gates, and cell-cell communication pathways. (c) Recombinant genes are used to turn host organisms into manufacturing systems for products such as biofuels, pharmaceuticals, and protein therapeutics. (d) Many challenges exist in improving robustness of engineered strains. Synthetic genes compete for resources with existing cellular processes for survival and replication. Host organisms will find ways to mutate away or repress synthetic pathways that are too resource expensive or deleterious to growth. Photographs in part (a) are from Wikimedia Commons - usage permitted under the Creative Commons license.

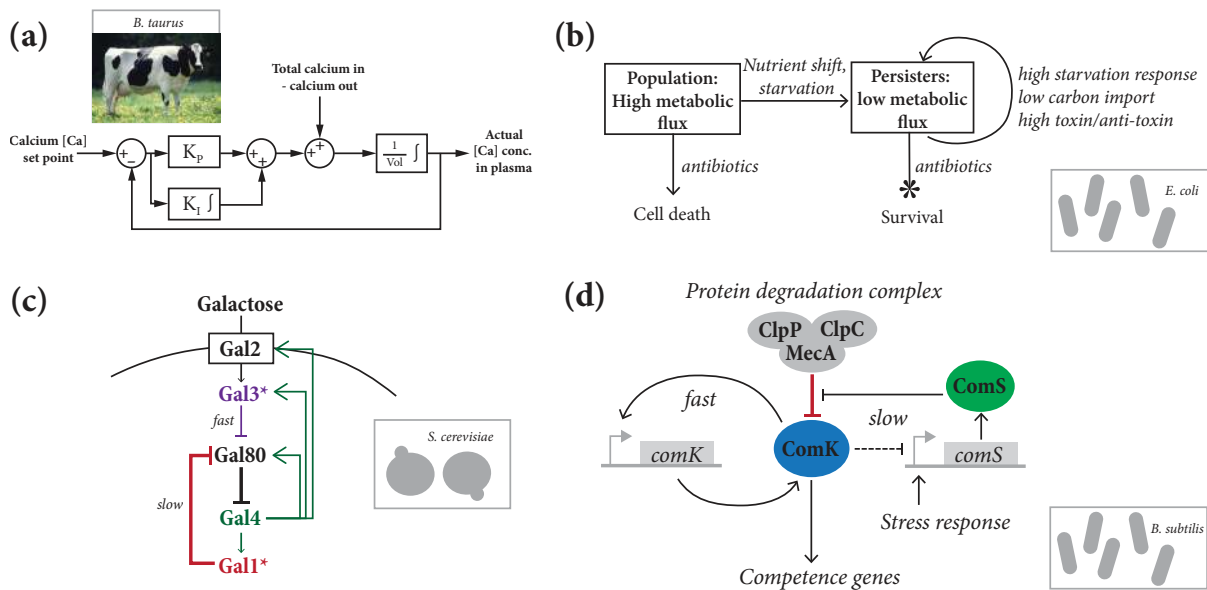


Figure 3. Examples of biological controllers in nature. (a) Proposed mechanism of PI control in cows to regulate the calcium concentration in plasma [40]. (b) Positive feedback as a possible example for maintenance of bacterial persisters within an *E. coli* population [45]. (c) Regulation of galactose metabolism in *S. cerevisiae* yeast involves layers of positive and negative regulation of the main repressor Gal80, resulting in bimodality and hysteresis [47]. (d) Regulation of single cell competence in the gram-positive bacterium *B. subtilis* features activation through competitive inhibition of protein degradation machinery [48].

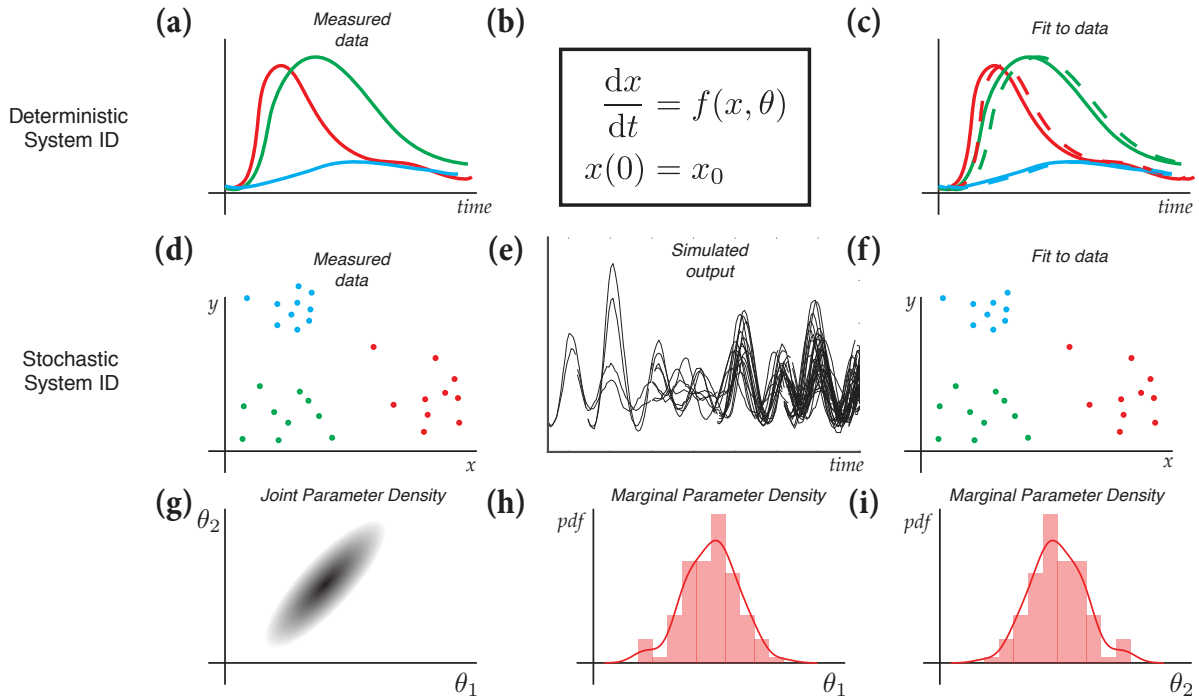


Figure 4. Bayesian system identification for deterministic and stochastic synthetic gene circuits. (a) Deterministic data is collected on a bulk culture over time. (b) A model using ODE's is used to model the synthetic gene circuit dynamics in a deterministic fashion. (c) The model fit is evaluated by matching simulated and measured trajectories. (d) Stochastic data is collected by measuring a distribution of outputs X and Y across a population of cells. (e) Stochastic simulations are used to compute the distribution of the output given a stochastic model. (f) The fit of the model to the data is evaluated by comparing output distributions across the population (g,h,i) The output of the procedure in both cases is a posterior density function over parameters. The joint parameter density function can show correlations between parameters, suggesting parameter degeneracy.

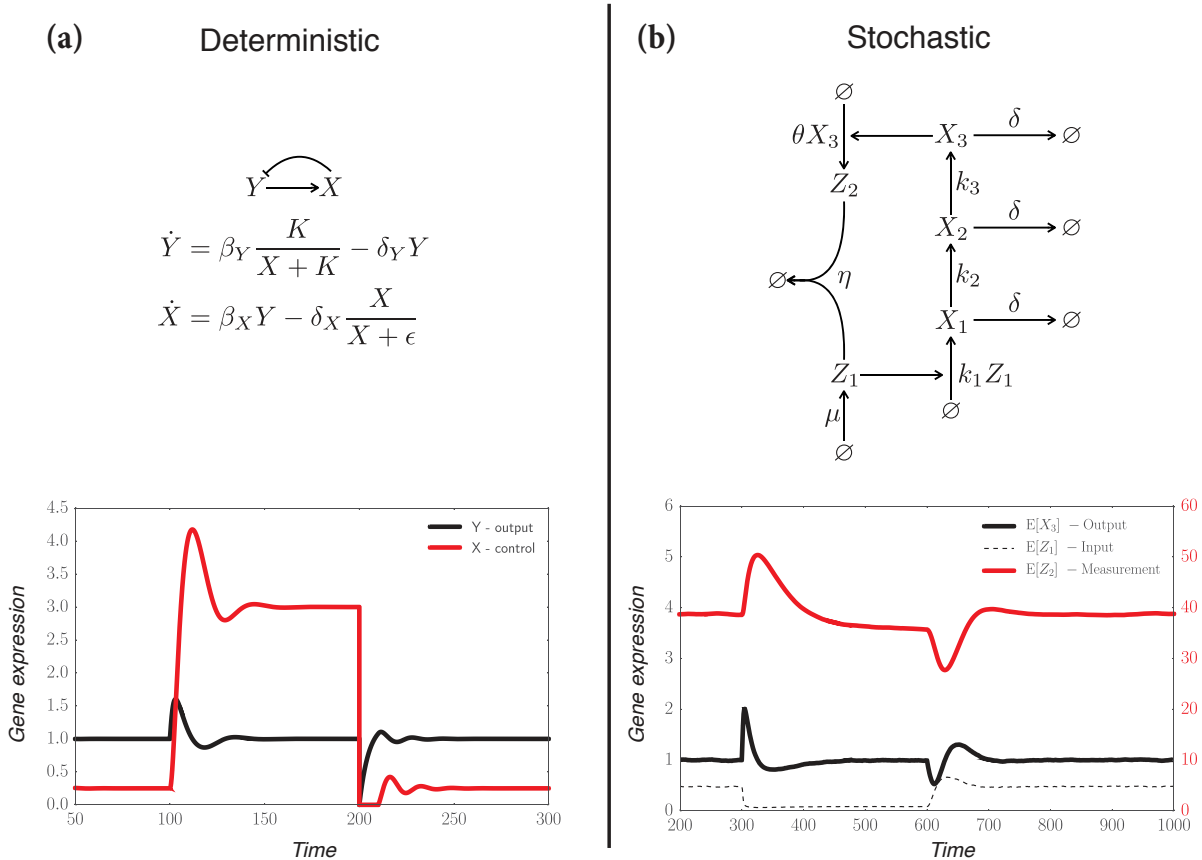


Figure 5. Integral control of gene expression in synthetic gene circuits. (a) In the deterministic case, a control species X , which represses Y , has dynamics dependent only on Y , ensuring exact adaptation of the steady state Y value. (b) In the stochastic case, the existing network consisting of species X_1 , X_2 , and X_3 is augmented with control species Z_1 (control input) and Z_2 (measurement). The added reactions involving Z_1 and Z_2 exactly control the mean output $E[X_3]$ to unity. The rates k_1 , k_2 , k_3 are sped up at $t = 300$ and then returned to their original values at $t = 600$. The output, despite being perturbed, always recovers to its desired value of 1. The plotted expected values are calculated across a set of 100,000 stochastic simulations.

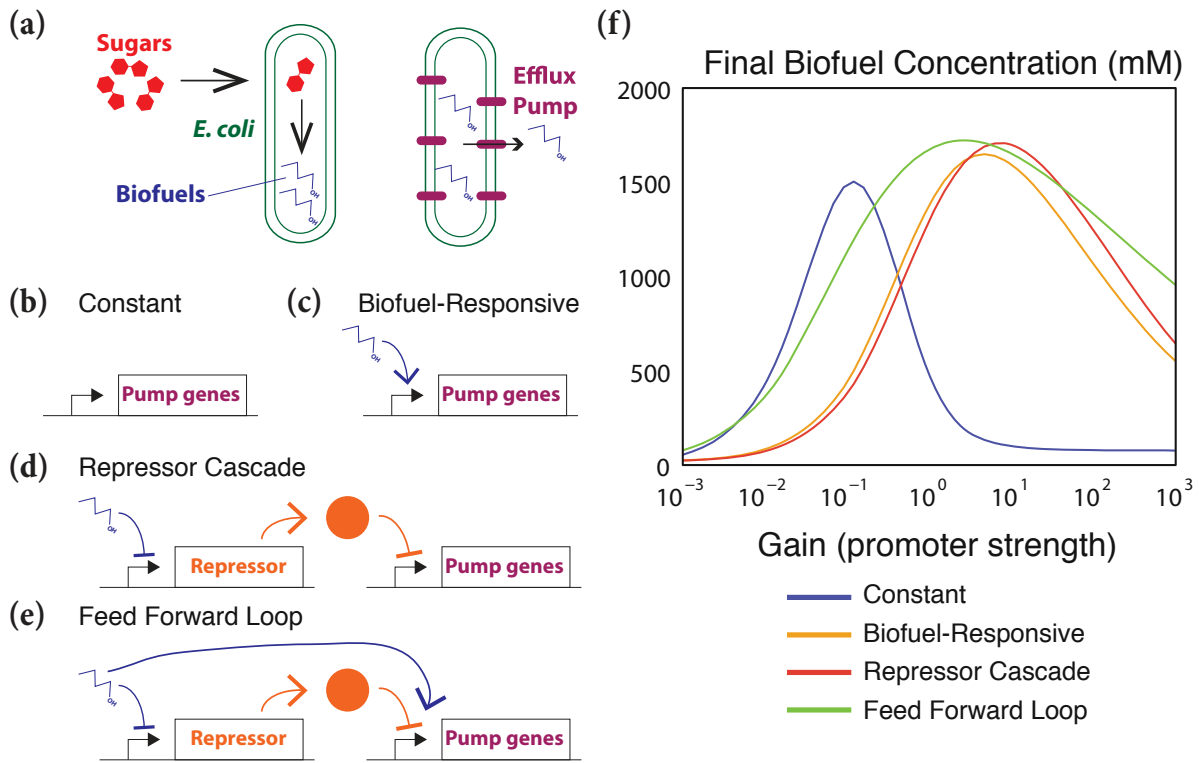


Figure 6. Control architectures for biofuel export [83]. (a) *E. coli* produce biofuels from sugar but must export the biofuels using efflux pumps to avoid toxicity. (b) Open loop expression of pump genes at a constant rate. (c) Closed loop expression of pump genes at a rate dependent on biofuel concentration. (d) Closed loop regulation of pump genes by repressing expression of a repressor that represses pump genes. (e) Closed loop regulation of pump genes by combining the two prior architectures. This is a feed-forward loop because biofuels not only immediately activate pump gene expression but also cause further activation by repressing the repressor. (f) Simulated final biofuel concentration as a function of the promoter strength (from Figure 6 in [83]). The feed-forward loop provides the most robust behavior, as the biofuel production is high across a wide range of promoter strengths. The open loop constant expression strategy only performs well in a narrow range of promoter strengths.

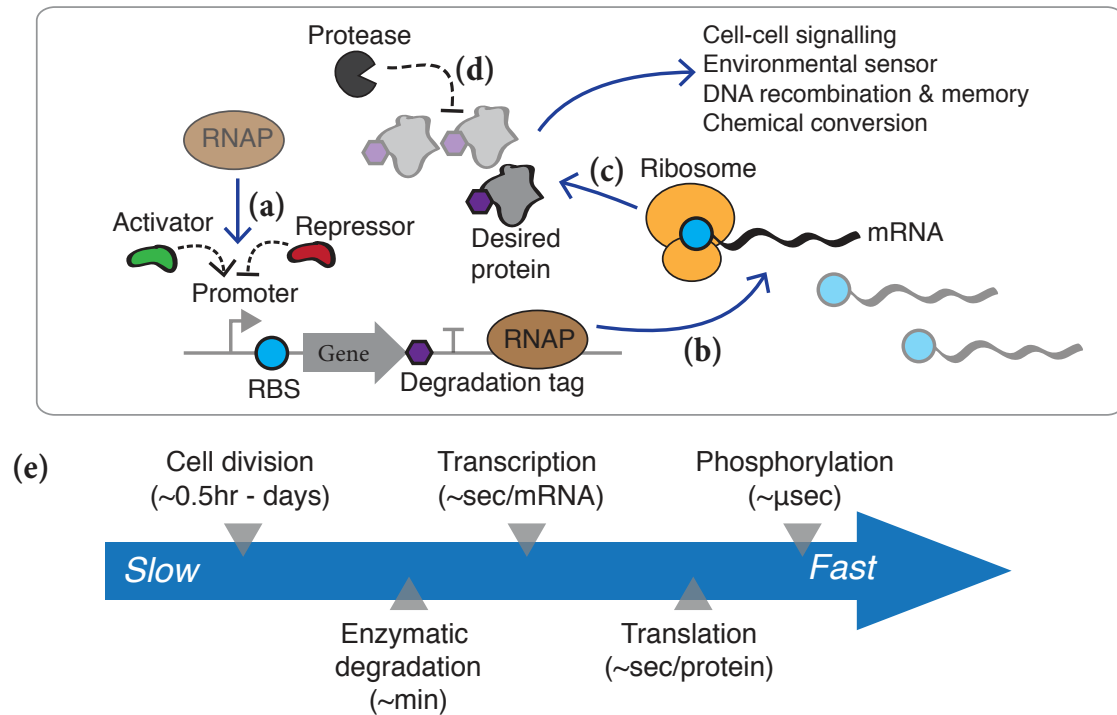


Figure 7. Molecular components and relative timescales. (a) Transcription of a desired gene can be regulated using activator and repressor proteins to either recruit or inhibit DNA polymerase. (b) DNA is transcribed into mRNA, which contains a ribosome binding site (RBS). RBSs can be engineered for different translation efficiencies, which affect ribosomal binding rates. (c) The ribosome translates the mRNA into sequences of amino acids, which then fold to form the desired protein product. (d) Proteins are constantly being degraded by housekeeping proteases. The addition of a degradation tag will specifically target the proteins to these proteases, ensuring much shorter half-lives for quicker turnover. (e) Relative timescales for the essential biological processes. Cell division can vary from 20 minutes for bacteria to several days for mammalian cells. Phosphorylation happens on the order of μ seconds, and is often only limited by the distance between proteins, rather than reaction time.

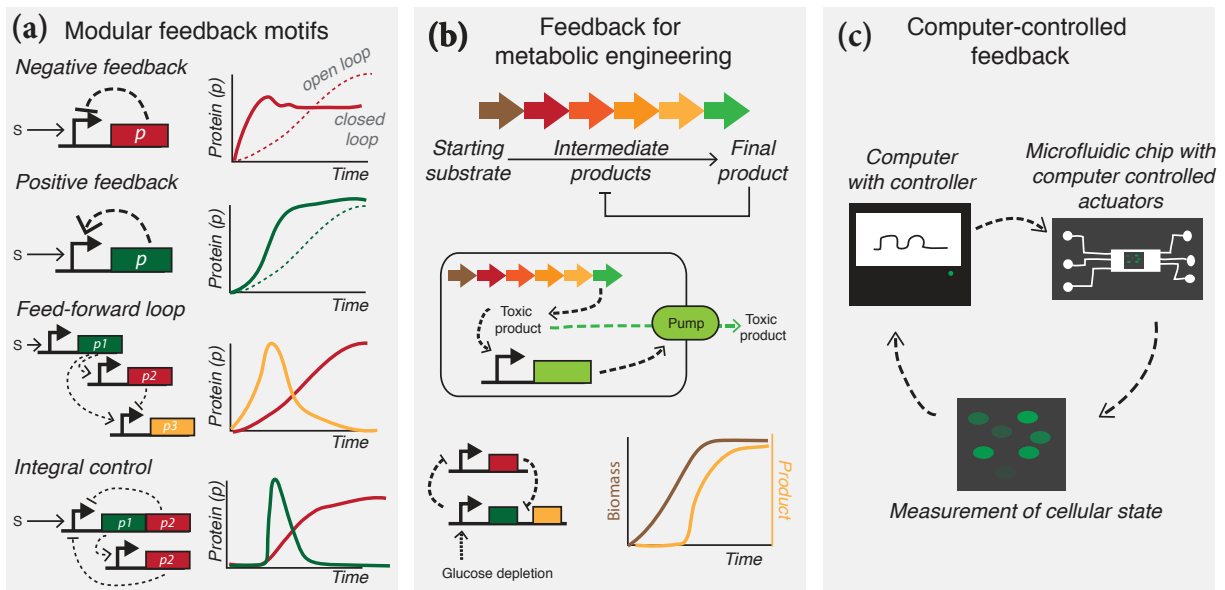


Figure 8. Implementation of feedback control in living cells. (a) Synthetic feedback modules have been built to demonstrate the feasibility of feedback in engineered organisms. (b) Feedback controllers in metabolic engineering are used to reduce toxicity and increase production. This can be achieved by sensing the final product and repressing overproduction intermediates, or product-dependent production of pumps to export molecules outside of the cell. Switches can also be utilized to separate biomass generation from production. (c) Computer-controlled feedback uses traditional controllers on sophisticated microfluidic actuators that control living cells.

Sidebar: Synthetic Biology in Context

4 Synthetic biology is a subfield of bioengineering that overlaps with the fields of systems
biology, metabolic engineering, and protein engineering (Figure S1). Most researchers in these
6 areas are associated with at least two of these categories. Synthetic biology might be most easily
contrasted with systems biology – systems biology seeks to understand natural networks while
8 synthetic biology seeks to design and build completely new networks. For a given biological
organism, systems biology attempts to elucidate the complex inner workings of natural pathways
10 for signaling and regulation. The systems biologist might introduce genetic modifications into the
organism in order to test hypotheses about molecular interactions or mechanisms of regulation.
12 This leads to understanding about the natural networks that govern cellular behavior. A synthetic
biologist would use these same genetic tools to design and build a completely novel molecular
14 pathway within the same organism. In characterizing the performance of these man-made
pathways, the synthetic biologist also learns about the natural networks that he or she is
16 competing with and works to minimize interference between these systems.

While synthetic biology broadly encompasses any research in which non-native genes
18 are used to build new systems, metabolic engineering refers specifically to research that seeks
to optimize production of a certain chemical or protein. Metabolic engineering focuses on
20 understanding and optimizing natural biosynthetic pathways such that high quantities of a desired
product, such as a biofuel, can be achieved when cells are grown in large quantities (usually
22 brewery-like fermentation tanks). These natural biosynthetic pathways often come from plants
and contain at least 10 – 20 genes. Introducing them into a bacterial or yeast host requires a
24 great deal of optimization and understanding of how molecular are converted down the pathway.
Most of the current commercial applications of synthetic biology are examples of metabolic
26 engineering.

All biomolecular pathways are actuated by proteins – proteins can bind specifically to
28 DNA and RNA, to other proteins, and to chemicals and metabolites within the cell. Once bound
to their target, proteins carry out a huge variety of specific functions such as DNA repair,
30 chemical conversion, and signal transduction. Although it is known how the DNA sequence in
a gene is translated into a sequence of amino acids (see Sidebar “Central Dogma”), predicting
32 how these amino acids actually fold in three-dimensional space into a functional protein is
very difficult and computationally intensive. Engineering proteins to have new functions or to
34 bind to new targets requires an understanding of the 3D structure, and often involves screening
large libraries of protein mutants for the desired characteristics. The field of protein engineering
focuses on predicting, designing, and creating new protein variants through these methods. These
new proteins can then be used throughout synthetic biology, systems biology, and metabolic

2 engineering.

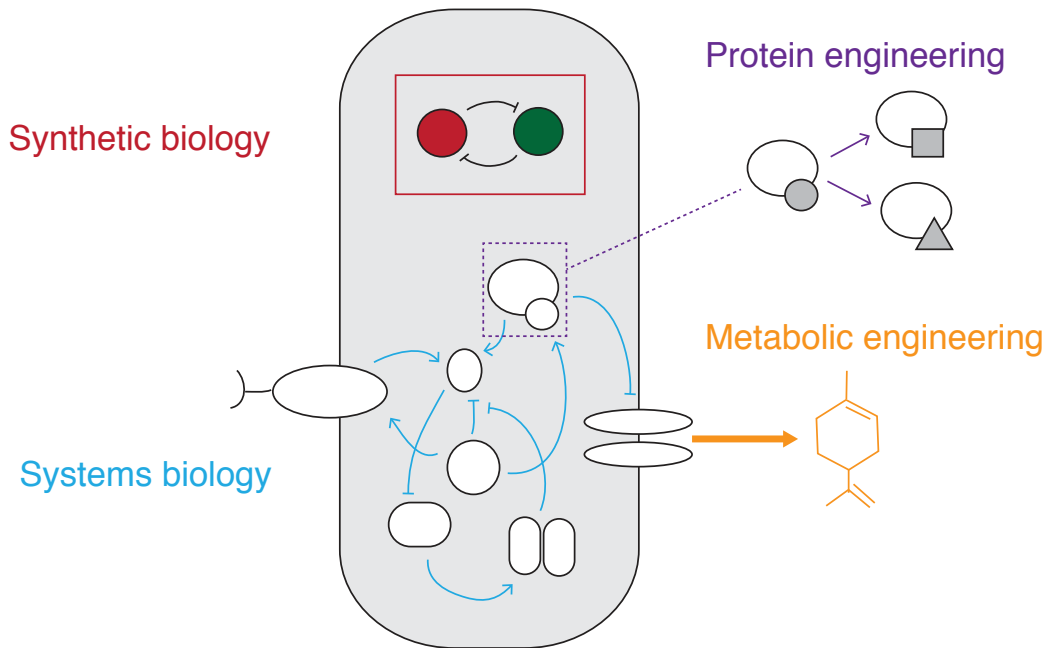


Figure S1. Sidebar: Synthetic Biology in Context. Synthetic biology overlaps with the fields of systems biology, protein engineering, and metabolic engineering. While synthetic biology refers broadly to any use of genetic engineering to design and build non-native biological systems, systems biology aims to study existing pathways and signaling networks within an organism. Metabolic engineering refers to the genetic engineering of organisms to produce a specific chemical or biological product, and protein engineering focuses on understanding and engineering the structure and function of individual proteins.

Sidebar: Central Dogma

4 All biological organisms rely on the storage of information in DNA. DNA is a polymer
in which each monomer is one of four nucleotides: adenine (A), guanine (G), thymine (T), or
6 cytosine (C). The sequence of A's, G's, C's, and T's in the DNA carries the necessary information
to produce proteins.

8 The RNA polymerase enzyme produces a messenger RNA (mRNA) molecule with an
equivalent sequence as the DNA in a process called transcription. Then, a protein and RNA
10 complex called a ribosome reads the mRNA sequence and produces a chain of amino acids in
a process called translation. The ribosome translates the mRNA in increments of 3 bases at a
12 time. Each 3 base unit is called a codon, and each codon is specific to one of the 20 amino
acids.

14 The final chain of typically hundreds of amino acids then folds and matures into a protein.
Proteins make up the machinery that performs most of the essential tasks in the cell. For
16 example, RNA polymerase is a protein, and the ribosome contains many different proteins.
Green fluorescent protein (GFP) is a protein commonly used in synthetic biology as a fluorescent
2 readout of gene expression.

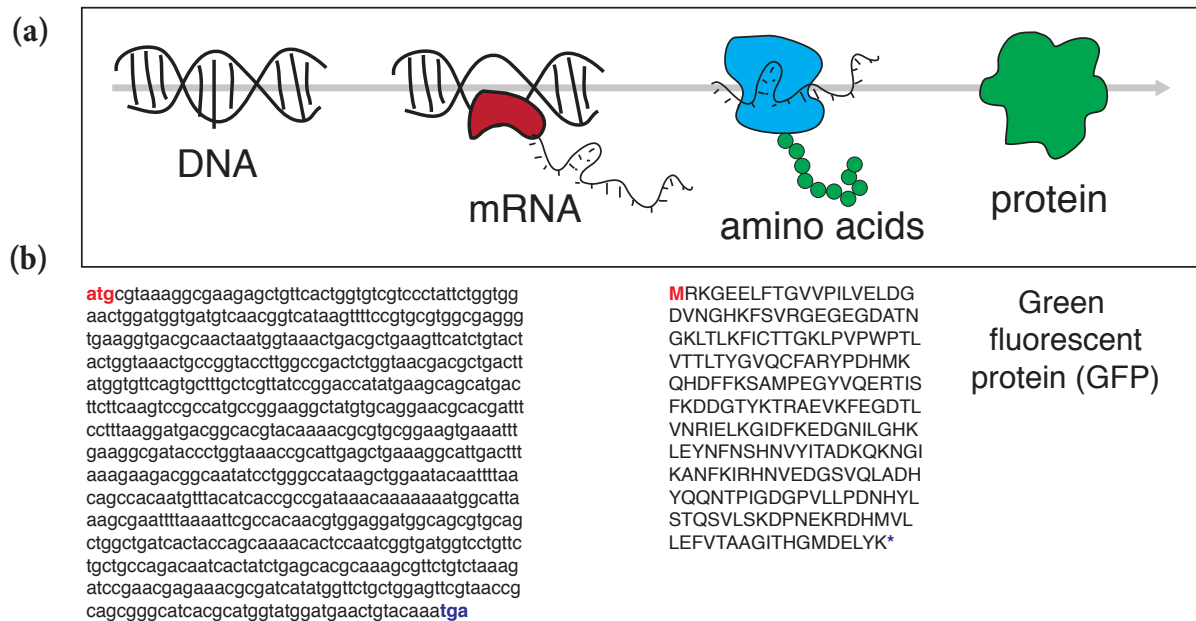


Figure S2. Sidebar: Central Dogma. (a) The DNA double helix is opened by RNA polymerase (red), which transcribes the DNA to mRNA. The ribosome (blue) translates the mRNA to an amino acid chain, which folds and matures to a protein (green). (b) The DNA sequence is a sequence of A,G,C, and T corresponding to the 4 nucleotides of adenine, guanine, thymine, and cytosine that make up the DNA polymer. The protein sequence is a list of the amino acids that make up the protein, where each letter corresponds to a different amino acid. There are 20 standard amino acids.

Sidebar: A Molecular Biology Workflow

4 Much of synthetic biology is still done at a small scale, particularly in academic labs. Here,
a typical cloning workflow is illustrated. In cloning, the basic goal is to isolate a desired gene
6 and insert it into a desired plasmid “vector”. Various plasmid vectors differ in properties such
as antibiotic resistance, copy number per cell, and host organism compatibility. At the scale of
8 a single graduate student or postdoc, this process can take three days on average.

In the first step of cloning, the target gene is isolated from its original source (genomic
10 DNA, another plasmid) via polymerase chain reaction (PCR), a process in which DNA can be
exponentially amplified through repeated cycles of heating and cooling. PCR requires the design
12 and synthesis of short DNA primer fragments (~16-20bp), which can be ordered commercially
and have a turnaround time of about 24 hours. The primers and original source DNA are
14 added into a microliter (uL) scale buffered solution containing DNA polymerase and nucleotides.
Modern polymerase enzymes can amplify 1 kilobase in 15 seconds, and typical PCR protocols
16 can be completed in one hour. The linear PCR products are then run on an electrophoresis gel
with a DNA ladder and inspected for size.

18 Following purification of the PCR product, the gene is then considered “linearized” and
sufficiently amplified to be inserted into the destination vector. Modern scientists have a variety
20 of options for combining gene and vector (*e.g* Gibson assembly, Golden gate assembly, and
variations thereof). These protocols utilize enzymes with unique DNA recognition and cutting
22 properties, and all take about 1 – 2 hours to complete with 5 – 10 uL reaction volumes.

In order to select for correctly assembled plasmids, specially prepared *E. coli* are
24 transformed with the assembly reaction. These *E. coli* have been prepared to increase likelihood
of foreign plasmid uptake, a process known as transformation. In this process, it is assumed that
26 only a small percentage of the cells will receive a plasmid, and those that do will only have one.
The *E. coli* are then grown on nutrient-rich agar plates with selective antibiotic that matches the
28 resistance of the plasmid vector, thus ensuring that only cells with correct assemblies survive
and propagate. The growth of cells on the selective agar requires 12 – 16 hours at 37C.

30 The surviving cells will appear as monoclonal populations on the selective agar – these
monoclonal populations are known as colonies. Within each colony all of the cells are
32 genetically identical but between colonies, there may be single base mutations or other errors.
Therefore, several colonies are usually prepared for sequencing confirmation. The process of
34 preparing the DNA and submission to commercial sequencing entities usually takes about 24
hours. Once a newly constructed plasmid has been sequenced verified, it can be used for
2 experiments.

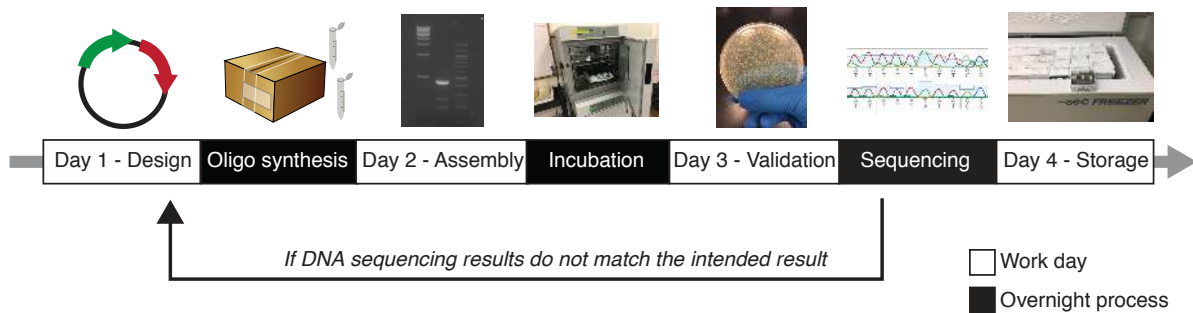


Figure S3. Sidebar: A Molecular Biology Workflow. The DNA construct is designed (Day 1), and the required DNA oligonucleotides are ordered for shipment the following day (Oligo synthesis). The DNA is then assembled using laboratory techniques such as restriction digest cloning, Gibson assembly, or Golden gate assembly. Common techniques such as gel electrophoresis are used to image DNA (Day 2, inset). Bacteria carrying the desired construct are incubated overnight on agarose with selective antibiotics – this only allows cells with desired DNA assemblies to grow. Single colonies of bacteria can then be observed the next day (Day 3). The DNA from the colonies is sent in for DNA sequencing (Sequencing), and the correct construct can be stored indefinitely in a bacterial culture at -80C (Day 4).

Sidebar: Uncertainty in Biology

4 A key differentiating feature of biological systems from those in traditional engineering
disciplines is the incredible amount of uncertainty that can affect the performance of a synthetic
6 gene circuit. Synthetic gene circuits do not operate in isolation, yet our understanding of the
“black box” that is the host cell is very limited. Living organisms already have complex networks
8 of feedback regulation and redundancy; these networks are still actively being elucidated by
researchers. The phrase “context dependence” is used to describe the effects of unknown or
10 dynamically changing surroundings on the performance of a synthetic gene circuit. There are
many different types of context dependence effects. For example, within the cell a synthetic
12 circuit may compete with other genes for resources, and its performance may depend on the
growth rate of the organism or its internal metabolic state. Additionally, any synthetic circuit relies
14 on global gene expression machinery such as polymerases and ribosomes, and these components
can vary between different species of organisms, different strains of the same species, or even
16 different cells of the same strain.

At the level of the environment, synthetic gene circuits can be affected by factors such
18 as pH, temperature, oxygen availability, and other factors. Often, these factors do not affect
the synthetic gene circuit directly, but instead they affect the internal state of the organism,
20 which then affects the synthetic circuit. Many of these environmental factors trigger internal
feedback systems and stress responses that allow the cell to survive in harsh conditions. For a
22 comprehensive review of context dependence effects in biology along with references describing
the context dependence effects mentioned here and others, [130] is an excellent resource.

24 Robust control theory is often used in control engineering to mitigate the effects of
uncertainty in the dynamics of a system while maintaining some level of control performance.
26 The trade-off is that the control system will be robust with respect to some uncertainties but
fragile with respect to others. This methodology typically works best when the engineer has
28 some intuition as to which possible sources of uncertainties are likely to be significant, and
which sources of uncertainties can be safely ignored. This intuition is lacking for synthetic
30 gene circuits. In fact, given a specific gene circuit in a specific strain of cells operating at
a specific set of conditions, it is typically unclear prior to doing experiments which possible
32 context dependence effects will significantly affect the system.

Researchers in synthetic biology take a number of approaches to mitigate these unknowns.
34 Usually, this involves reducing as much variability as possible in experimental conditions (e.g.
using a well studied host organism and sticking to a single strain, strict adherence to growth
2 conditions and media formulation, carefully maintaining frozen stocks of engineered strains

with reproducible results) and then doing extensive characterization of circuit performance in
4 those conditions. These measurements, while limited by the types of data that can be collected
("Sidebar: Measurement Types"), allow for some modeling parameters to be grounded in reality,
6 and even predictive for that particular experimental setup. The stringent conditions from which
modeling parameters are often collected may be challenging for creating broadly applicable
2 models, and even more so for designing robust controllers.

Sidebar: Measurement Types

4 There are three main types of data that can be collected on synthetic gene circuits operating
6 *in vivo*: bulk data, single cell population snapshot data, and time series single cell data. These
8 measurement types and the associated output data are illustrated in Figure S4. Bulk data consist
10 of measurements of an output for a culture of cells growing in a container, where the measured
12 output can be thought of as proportional to the sum of the outputs for all the cells in the
14 culture. Population snapshot measurements consist of measuring an output for a large number
of individual cells in a population. However, the same cells cannot be followed over time, and at
the next time point, a different set of cells will be measured. Population snapshot measurements
include techniques such as flow cytometry and mRNA FISH (fluorescence *in situ* hybridization).
Finally, with time series measurements, a single cell's output can be tracked over time using
techniques such as time lapse fluorescence microscopy.

Automation heavily aids data collection for synthetic gene circuits. For bulk data, plate
16 reader machines can incubate cell cultures at a specific temperature and collect culture density
and fluorescence information at regular time intervals. This can be done for hundreds of wells
18 on a single plate in parallel. In the case of single cells, microscopy can also be automated. Using
automatic stages, autofocus capability, and the appropriate software, images can be acquired for
20 over ten positions in parallel at regular time intervals. Finally, flow cytometers can process a
cell culture and read both cell size and fluorescence for tens of thousands of cells within a few
22 seconds.

It is important to note that each of these measurement types provides progressively more
24 information than the previous type. If we ignore the time series relationships in time series data,
then it can be considered as a set of population snapshots at different time points. If we take the
26 mean or sum of a population snapshot output distribution, we can think of the resulting value
as a bulk measurement.

28 Finally, the type of measurement collected can and should impact the choice of model for
the system being studied. As an example, if the data collected is all bulk data, then using a
stochastic model that accounts for noisy gene expression in single cells could be unnecessarily
2 complex, and a deterministic model based on ODE's might be a better choice.

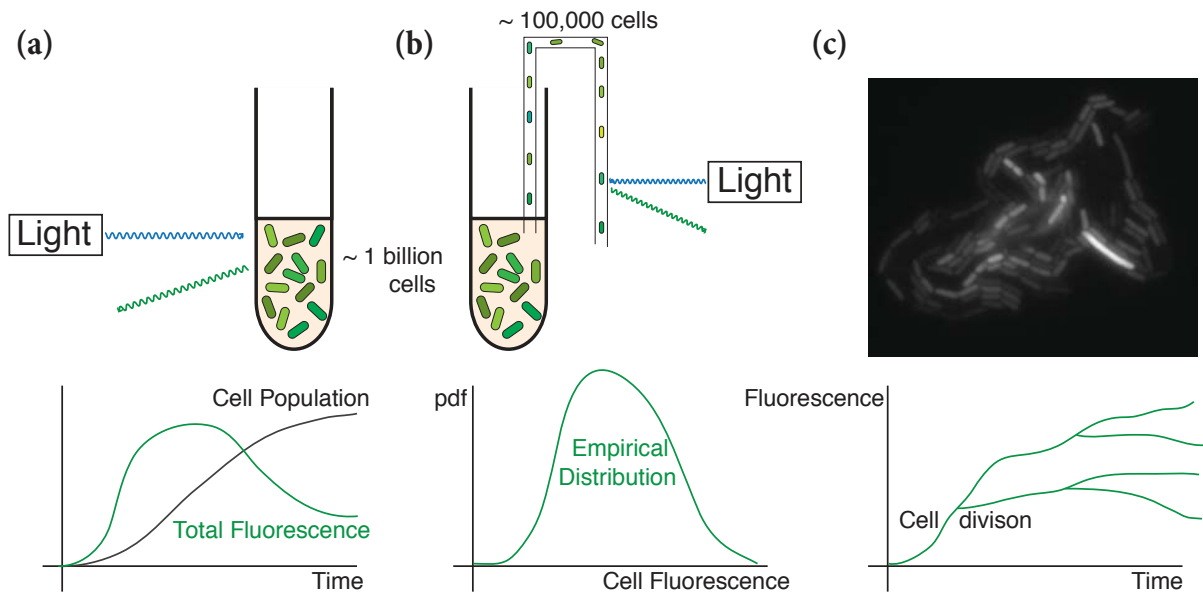


Figure S4. Sidebar: Measurement Types. The different types of measurements for synthetic gene circuits. (a) In bulk measurements, total fluorescence and cell population (by measuring optical density) can be measured for a culture over time using a plate reader. (b) In flow cytometry, single cells pass through a detector that measures fluorescence for many individual cells. (c) In time lapse microscopy, fluorescence output can be tracked across a growing lineage of single cells.

4 Sidebar: Model Classes

4 Synthetic gene networks are typically modeled as systems of chemical reactions occurring
inside a cell where the chemical species in the circuit, proteins and mRNA's, are produced and
6 degraded with different reaction rates. The level of one species can affect the rate of production
or degradation of another species. This type of interaction is what gives rise to circuit type
8 behavior.

For example, in Figure S5a, a simple synthetic gene circuit where two genes X and
10 Y repress each other's production can give rise to a bistable toggle switch. In models of
gene circuits, it is also common to ignore mRNA production and degradation as mRNA is
12 an intermediary between DNA and protein that typically has fast dynamics and no function of
its own. In the model of the toggle switch in Figure S5b, the mRNA dynamics are ignored, and
14 the model simply has the two protein levels X and Y repressing each other's production.

The systems of chemical reactions can be modeled deterministically or stochastically.
16 Deterministic models consist of a set of ODE's (see Figure S5b). In the stochastic case, each
reaction is modeled as a discrete event that occurs with a certain propensity [131]. This results
18 in a continuous time discrete state Markov chain whose probability evolution is described by
the chemical master equation [132].

20 Simulation trajectories of the deterministic system can be generated using a standard
ODE solver. For stochastic models, sample paths of the system can be generated using a
22 stochastic simulation algorithm [131]. Also, the time evolution of the full state distribution
can be approximately computed by using techniques such as moment closure [133] or finite
24 state projection [57]. In moment closure, a system of ODE's approximately describes the time
evolution of the moments of the state distribution. In moment closure, the state space is truncated,
26 so that the time evolution of the state distribution can be approximated by solving a high
dimensional linear system of ODE's.

28 These methods all describe ways to model synthetic gene circuits as chemical reactions
inside a single cell. However, some synthetic gene circuits operate at a multicellular level, and
30 thus, cell growth and death must be modeled. In a deterministic setting, cell growth is typically
modeled as logistic growth with a saturating carrying capacity that depends on the type of cells
32 and their environment. In the stochastic case, it is important to note that when a cell divides,
its contents must be partitioned between its daughter cells. The randomness of this partitioning
can introduce further noise into gene circuit dynamics.

2 When matching synthetic gene circuit models to experimental data, it is necessary to choose

a model appropriate for the type of data being collected. For example, if bulk population average
4 data is being collected, then a deterministic model might be more appropriate than a stochastic
model. It can also be necessary to iterate on the reactions themselves to get the best fit to the
6 data. A number of software packages exist that assist in modeling, simulation, and parameter
inference of synthetic gene circuits [134], [72], [135], [136].

8 A deterministic and a stochastic simulation of the toggle switch circuit from Figure S5a
are presented in Figure S5d. The deterministic simulation stays at a low X and high Y steady
2 state, while the stochastic simulation can randomly switch to the opposite state due to noise.

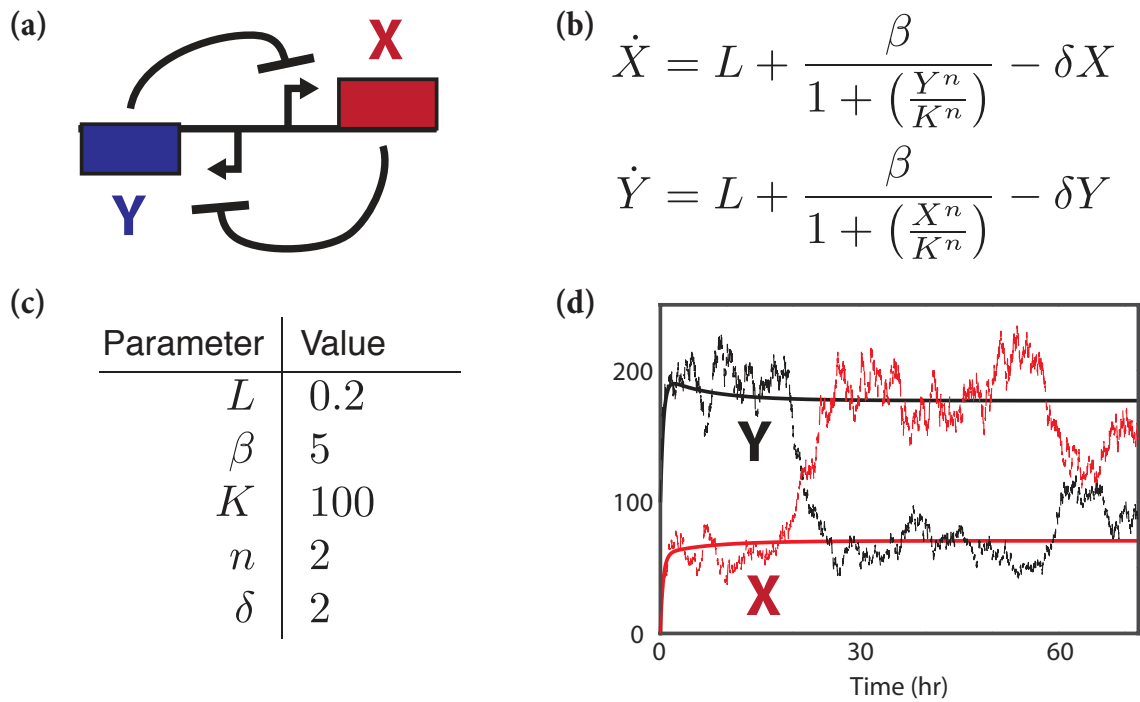


Figure S5. Sidebar: Model Classes. Modeling of a toggle switch synthetic gene circuit. (a) In a two species toggle switch, two genes X and Y repress each other's expression, which creates two steady states where either X is high and Y is low or vice versa. (b,c) Deterministic ODE's and parameters for a toggle switch model. Ignoring mRNA, X and Y are produced at a basal level with additional production repressed by the other gene along with first order dilution/degradation. (d) A deterministic (solid) and stochastic (dashed) simulation of the toggle switch. The deterministic simulation stays at a Y high/ X low steady state while the stochastic simulation can switch due to noise.

Sidebar: Integral Control for Synthetic Gene Circuits

4 Here, we cover the details of Figure 5, which shows both a deterministic [79], [80] and stochastic [82] version of integral control for synthetic gene circuits.

6 In Figure 5a, the output to control is Y and the control input is X . We have a system of ODE's describing the dynamics

$$\begin{aligned}\dot{Y} &= \beta_Y \frac{K}{K + X} - \delta_Y Y \\ \dot{X} &= \beta_X Y - \delta_X \frac{X}{X + \epsilon}.\end{aligned}$$

8 In this system, X represses production of Y , and Y activates production of X , which creates a negative feedback loop from Y to itself. Additionally, Y is also subject to a dilution
10 term. We assume that X is actively degraded by an enzyme much more quickly than the dilution rate, so we ignore dilution for X . Furthermore, we assume that the degradation of X saturates at
12 a low level of X , so $\epsilon \ll 1$. This allows us to approximate the second equation as $\dot{X} = \beta_X Y - \delta_X$. This means that the steady state value of Y will always be $\frac{\delta_X}{\beta_X}$, regardless of the values of β_Y and
14 δ_Y . The plot in Figure 5a shows the levels of X and Y as β_Y is varied. When β_Y is increased, Y production initially increases, but then X also increases and represses Y back to the same steady
16 state. When β_Y is lowered back to the original value, the original steady state is recovered.

In Figure 5b, we have a stochastic system of chemical reactions in which producing an
18 actuator species Z_1 leads to eventual production of output species X_3 through some chemical reactions. In this case, the set of chemical reactions is a cascade, but in general, the chemical
20 reactions can be arbitrary. The output X_3 leads to production of a sensor species Z_2 , which can sequester and annihilate the actuator species Z_1 . We can write down a subset of the moment
22 equations for this set of reactions.

$$\begin{aligned}\frac{d E [Z_1]}{dt} &= \mu - \eta E [Z_1 Z_2] \\ \frac{d E [Z_2]}{dt} &= \theta E [X_3] - \eta E [Z_1 Z_2]\end{aligned}$$

At steady state, $E [X_3] = \frac{\mu}{\theta}$. This allows the control designer to set the expected value of
24 the output independent of the rates within the actual set of chemical reactions. In Figure 5b, it is shown that changing the rates k_1 , k_2 , and k_3 does not affect the expected value of the output
2 at steady state.

Sidebar: Attenuation of Loading Effects in Synthetic Gene Circuits

4 Synthetic gene circuits often run into the types of impedance effects that are common in
electrical circuits. In electrical circuits, an operational amplifier is often used to introduce a high
6 gain that allows an output voltage to track an input voltage very closely. This is illustrated in
Figure S6a-b. The use of the op amp makes the steady state output voltage V_{out} less susceptible
8 to impedance effects that could be introduced by additional downstream circuit components.

In synthetic gene circuits, when a transcription factor is used to activate gene expression,
10 the transcription factor must bind to a promoter for a gene to recruit RNA polymerase. However,
as more and more downstream binding targets are added, more transcription factor will bind to
12 target sites, and the concentration of free transcription factor will decrease. Thus, adding more
downstream targets decreases the amount of free transcription factor, which decreases activation
14 strength. In order to mitigate this loading effect, one can implement a high gain feedback loop
that maintains the concentration of free transcription factor.

16 The system in Figure S6c is a system that regulates the concentration of free transcription
factor to mitigate loading effects. In this system, a transcription factor X can be phosphorylated
18 by an input U to form an active transcription factor X_P , which can bind to downstream promoters
and activate gene expression. However, if there are more downstream targets, more X_P is
20 sequestered into transcription factor-DNA complexes C , and the amount of free X_P decreases,
which would reduce activation strength of the target genes.

22 This system uses high gain negative feedback to resolve this loading effect. If there is
enough Y and enough total X in the system, as the free X_P decreases, the system can just
24 phosphorylate more X to replace the lost X_P . If X_P builds up to too high of a level, then the
phosphatase Y can quickly convert the active X_P back to its inactive form X .

26 The model for this system is given in Figure S6d. The two differential equations provide
a model of the dynamics in terms of active X_P and complexes C of X_P bound to downstream
28 target promoters. Under high gain conditions when the total amount of X and Y are sufficiently
high, at steady state the free X_P in the system is not affected by the load in the system p_{tot} ,
where p_{tot} is the number of downstream DNA binding sites. For a more detailed discussion of
2 this model and its load buffering properties, please see [127].

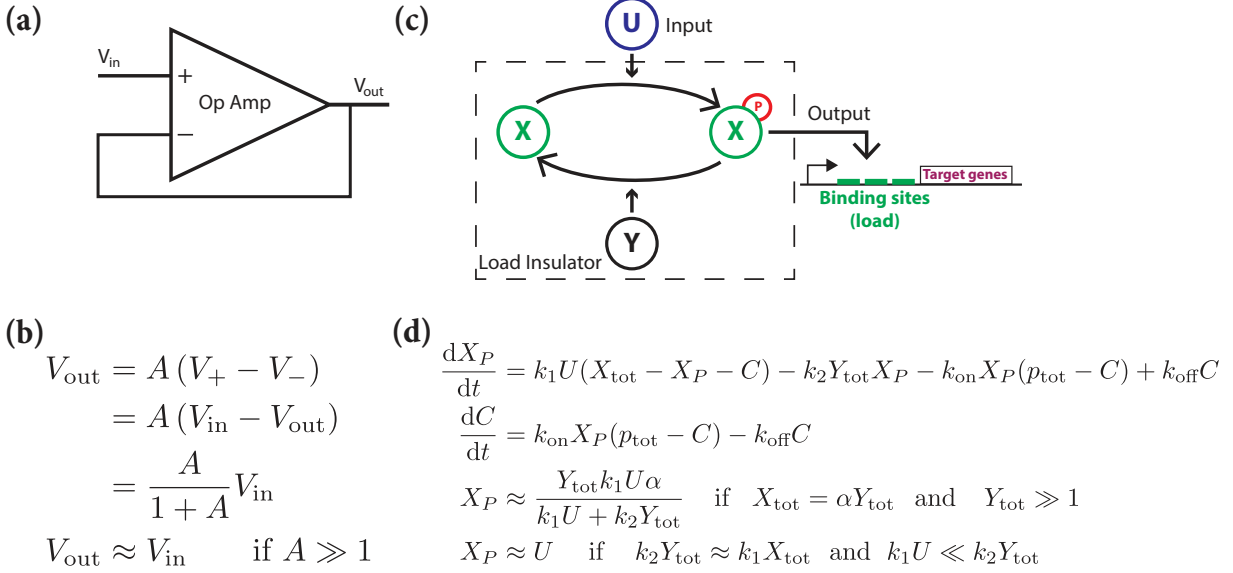


Figure S6. Sidebar: Attenuation of Loading Effects in Synthetic Gene Circuits. Attenuation of loading effects in engineered systems. (a,b) In an operational amplifier (op amp), high gain is used to ensure the output voltage V_{out} tracks the input voltage V_{in} even with impedance effects from downstream circuit components. (c) A load buffering device in synthetic biology [127]. The input U phosphorylates a transcription factor X , creating active transcription factor X_P . X_P can bind to downstream promoters and activate target gene expression, but this binding creates a load on the system. (d) The dynamics for the load buffering device written in terms of unbound active transcription factor X_P and bound promoter- X_P complexes C . The output X_P tracks the input U independent of load p_{tot} when certain high gain conditions are met.

Sidebar: Model Predictive Control of Gene Expression

4 This sidebar illustrates a simplified version of the type of model predictive control done
in [95]. The goal is to control the expression of a fluorescent protein in a cell culture to a
6 desired reference value. In this system, the cells are engineered to produce fluorescent protein
upon experiencing osmotic stress. The cells can be exposed to regular media, which means they
8 will not produce fluorescent protein, or they can be exposed to a high salt media that induces
osmotic stress, which leads to fluorescent protein production. This means that the input in this
10 system is binary (i.e. $u(t) = 0$ or 1), because the cells can either be subjected to osmotic stress
or not. Furthermore, the cells have a limited tolerance for osmotic stress, so the high salt media
12 must be applied in short pulses with sufficient time periods between pulses to allow the cells to
recover.

14 Following the model in [95], we model gene expression as a function of input with the
following set of equations.

$$\begin{aligned} \dot{x}_1 &= u(t) - \delta_1 x_1 \\ \dot{x}_2 &= kx_1 - \delta_2 \frac{x_2}{x_2 + K} \end{aligned} \tag{S1}$$

16 Here, the variable x_1 represents the host response to osmotic stress. The effect of past
osmotic stresses decays exponentially over time. The production rate of the output x_2 is
18 proportional to the osmotic stress response x_1 , and the x_2 protein can also be degraded at a
maximum rate of δ_2 .

20 For simplicity, we set all parameters in the model to 1. We consider the problem of tracking
a constant reference value for x_2 . Starting from an initial condition at the origin, the goal is
22 to minimize the squared error between x_2 and the reference over a finite time window of 0 to
20 minutes. Furthermore, because the cells have a limited tolerance for osmotic stress and due
24 to experimental limitations, we limit pulse durations to be between 1 and 5 minutes, and we
require a minimum gap of 3 minutes between pulses. To simplify the computation, we allow a
26 maximum of three pulses in the time window. This yields a constrained optimization problem in
which the error between reference and simulated output is minimized by optimizing the pulse
28 start and end times. This optimization problem is non-convex and can be solved locally using
general optimization software.

The results of this procedure are shown in Figure S7, which shows the optimal input
2 trajectories as well as the simulated output for reference values of 0.5 and 1.0. While the mean

tracking works in both cases, the fluctuations around the reference value are unavoidable for
4 this choice of parameters, because the time scale of the expression dynamics is too fast to be
controlled with the input. This is because the effective frequency at which the input can be
6 applied is limited by the required time gap between pulses. This type of time scale related
limitation is a common one in biological control. Of course, in an actual implementation of
8 model predictive control, these trajectories would be re-optimized at each time step.

In [95], the experiments yield much better results than the ones shown in Figure S7.
Application of model predictive control typically requires system identification and online state
2 estimation as well. The work in [95] details the system identification and state estimation
procedures for this system in addition to other details.

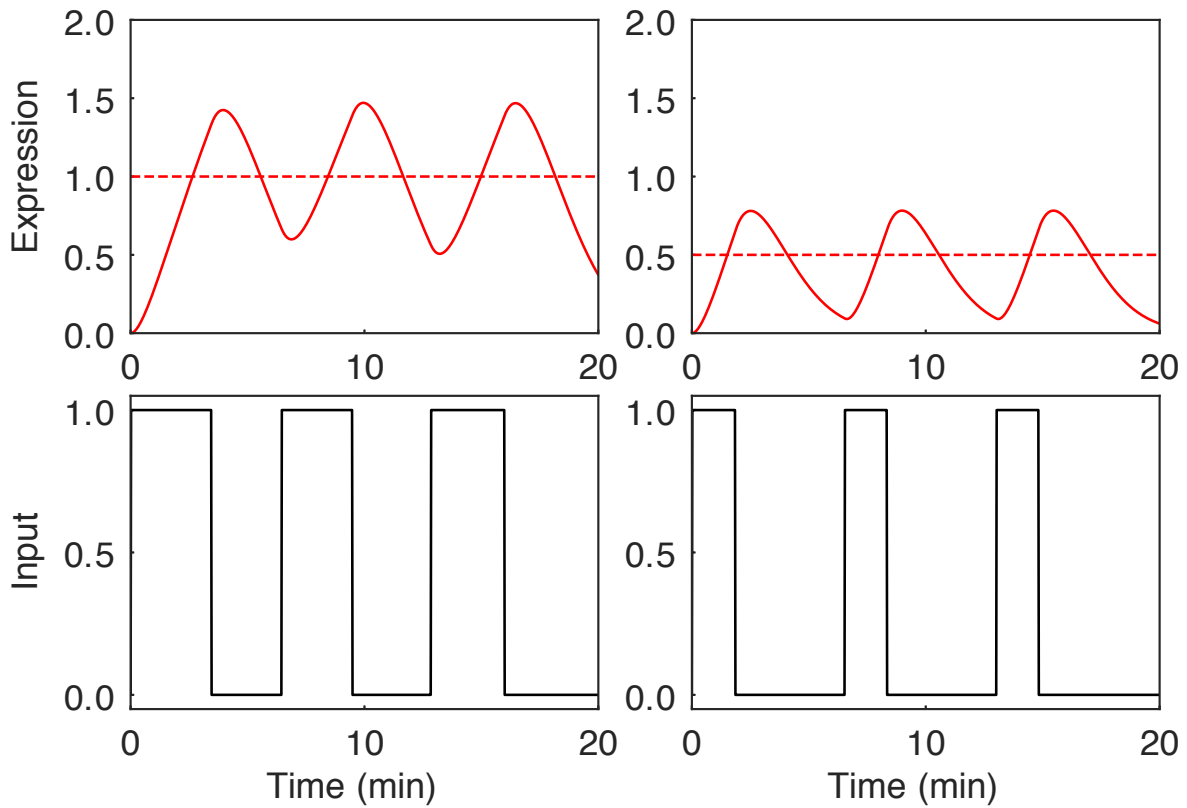


Figure S7. Simulated model predictive control of gene expression for the model given in (S1). Nonlinear optimization can be used to find optimal input pulses (black) for tracking a reference expression value (dashed red) with the simulated output (solid red). The procedure works for a reference of 1.0 (left) and 0.5 (right). Poor tracking performance highlights the potential time scale based limitations of biological control.

4 **Sidebar: Article Summary**

This review introduces basic concepts in synthetic biology and potential applications of control theory within this field. Synthetic biology is a subfield of bioengineering in which researchers use the tools of genetic engineering to build novel biomolecular pathways. In the last two decades, these synthetic pathways have been demonstrated in bacteria, yeast, and mammalian cells. These technologies potentially enable on-demand fermentation of necessary medicines, specialized organisms for bioremediation, or renewable production of molecules currently derived from petroleum. While many of these concepts have been proven in the laboratory, the field faces major challenges in reliability and robustness of engineered organisms outside of a very narrow controlled environment. These challenges arise because synthetic biomolecular systems must compete with natural processes for survival and replication. Living cells contain layers of complex, dynamic processes that interfere with even the simplest genetic engineering. This article describes basic concepts in synthetic biology, specific areas in which control theory could be applied, a general survey of the prior art, and the challenges with controlling and measuring the performance of biological systems. No prior background in biology or synthetic biology is assumed, and commonly used terms are defined for the reader.

Author Biography

4
Victoria Hsiao is a scientist at Amyris Inc. in Emeryville, CA. She received her B.S.
6 from the F.W. Olin College of Engineering in 2010 and her Ph.D. in Bioengineering from the
California Institute of Technology in 2016. Her interests are in the areas of synthetic gene circuit
8 design for feedback, detection, and memory in engineered organisms. Her current position as a
strain engineering scientist involves engineering yeast to produce molecules traditionally derived
10 from petroleum or non-sustainable sources.

Anandh Swaminathan is a postdoctoral scholar at the California Institute of Technology in
12 Pasadena, CA. He received his S.B. in Aerospace Engineering from the Massachusetts Institute
of Technology in 2011 and received a Ph.D. in Control and Dynamical Systems from Caltech in
14 2017. He is interested in integrating theory, computation, and experiments to improve the design
and implementation of synthetic gene circuits.

16 Richard M. Murray is the Thomas E. and Doris Everhart Professor of Control & Dynamical
Systems and Bioengineering at the California Institute of Technology in Pasadena, CA. He
18 received his B.S. from Caltech in 1985, and his Ph.D. from the University of California, Berkeley
in 1988. Research from his lab focuses on the application of feedback and control to networked
1606 systems, with applications in biology and autonomy. Current projects include novel control
system architectures, biomolecular feedback systems and networked control systems.