ACS SyntheticBiology

pubs.acs.org/synthbio Letter

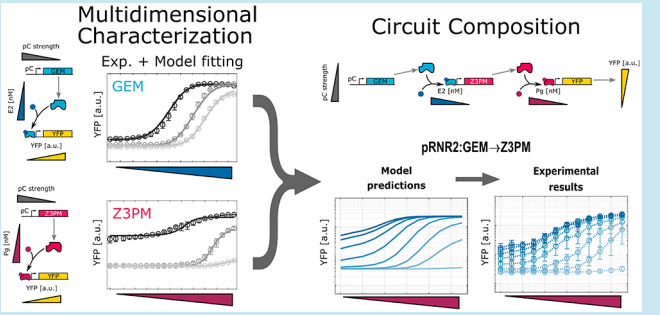
Multidimensional Characterization of Parts Enhances Modeling Accuracy in Genetic Circuits

Mariana Gómez-Schiavon,[#] Galen Dods,[#] Hana El-Samad,[*] and Andrew H. Ng[*]

Cite This: https://dx.doi.org/10.1021/acssynbio.0c00288 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Mathematical models can aid the design of genetic circuits, but may yield inaccurate results if individual parts are not modeled at the appropriate resolution. To illustrate the importance of this concept, we study transcriptional cascades consisting of two inducible synthetic transcription factors connected in series. Despite the simplicity of this design, we find that accurate prediction of circuit behavior requires mapping the dose responses of each circuit component along the dimensions of both its expression level and its inducer concentration. Using this multidimensional characterization, we were able to computationally explore the behavior of 16 different circuit designs. We experimentally verified a subset of these predictions and found substantial agreement. This method of biological part characterization enables the use of models to identify (un)desired circuit behaviors prior to experimental implementation, thus shortening the design−build−test cycle for more complex circuits.



KEYWORDS: *context, part characterization, genetic circuits, synthetic biology*

Synthetic biology utilizes biological parts such as transcription factors to build circuits that perform useful signal processing functions.[1,2] Advancements in DNA synthesis technology have rapidly grown the library of biological parts, but the construction of predictably performing circuits has lagged behind.[3] This lag is due in large part to two factors. First, it is now faster to build new DNA constructs than it is to characterize them experimentally, leading to the creation of many poorly characterized biological parts.[4] Second, simple phenomenological models of individual parts often fail to predict the behavior of circuits composed of these parts, even in the absence of contextual effects[5] or retroactivity.[6] Building more useful mathematical models of biological parts would greatly facilitate the forward design of genetic circuits with predictable behavior.[7−12]

A common feature of genetic circuits is the use of inducible synthetic transcription factors (iSynTFs)[13−15] as facile input nodes that can activate downstream elements in a dose-responsive manner. In *Saccharomyces cerevisiae*, a common architecture for iSynTFs consists of a fusion of a DNA binding domain (DBD), human hormone receptor (HR), and activating domain (AD).[13,16−18] Absent their corresponding hormones, these iSynTFs are sequestered in the cytosol via interaction of the HR with Hsp90.[19,20] This interaction inhibits nuclear localization until hormone is added, enabling dose-responsive control of transcription from a cognate promoter. iSynTFs are an indispensable part of the synthetic biology toolbox. Circuits containing iSynTFs have been used to probe the behavior of synthetic degradation-based feedback,[21,22] investigate noise in transcription,[13] and study the topology of endogenous circuits.[17,23,24]

iSynTFs are commonly characterized via their hormone dose response for one expression level of the transcription factor, but this represents only one dimension of their functionality. Genetic circuits often perform computation by modulating the expression level of transcription factors in a network. Thus, accurate prediction of circuit behavior should be contingent on understanding the behavior of these inducible transcription factors as they change expression level within a circuit.

In this work, we sought to predict the behavior of a simple genetic circuit: a transcriptional cascade consisting of two iSynTFs in which the first iSynTF activates expression of a second iSynTF. We initially used a simple Hill model to fit the behavior of each iSynTF.[25,26] Although this model reproduced the hormone dose response of an iSynTF at a single expression level, it failed at different iSynTF expression levels. We found that two modifications were essential to overcome this challenge. First, we developed an expanded Hill model to account for changes in the basal activity and output saturation as a function of iSynTF expression level. Second, in addition to

Received: May 30, 2020

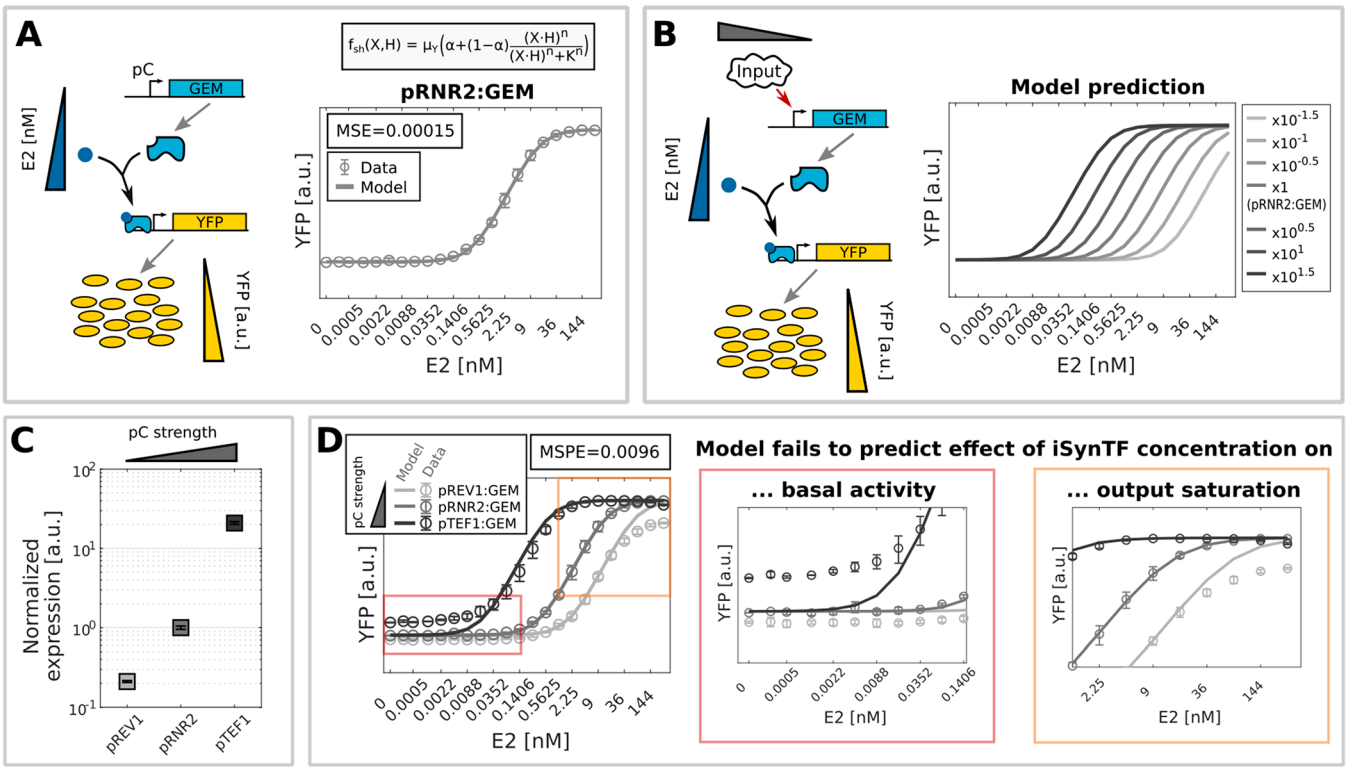


Figure 1. Simple Hill model fit to a single hormone dose response fails to capture the full behavior of iSynTF. (A) Left, A constitutively expressed (pC, constitutive promoter) inducible synthetic transcription factor (iSynTF; here GEM) is bound by its hormone inducer (here estradiol, E2) and activates transcription of a downstream YFP reporter. Right, Inducer dose response of GEM at a single expression level (pRNR2, constitutive promoter) as a function of hormone, here estradiol (E2). A simple Hill model (see inset: μ_Y, maximum synthesis rate; α, basal activity level; X, iSynTF concentration; H, hormone concentration; K, activation coefficient; n, Hill coefficient) was fit to the observed data (mean squared error, MSE = 0.00015). (B) Left: Expression level of GEM can change in response to inputs in a genetic circuit. Right: Simple Hill model prediction of inducer dose response for different expression levels of GEM (see legend for fold-change values). (C) Measurement of constitutive promoter expression levels using a pC:YFP fusion (where pC represents pREV1, pRNR2, or pTEF1). (D) Comparison of model predictions and experimental data for GEM inducer dose response at three different expression levels of GEM (mean squared prediction error, MSPE = 0.0096). Insets in red and orange boxes highlight the differences in basal activity and output saturation. Solid lines represent model predictions, open circles and filled squares represent experimental mean, and error bars represent s.d. of three biological replicates. See Supplementary Table S4 for used parameter values.

dependence on hormone concentration, we fit our model to experimental data collected from different expression levels of each iSynTF. Neither of these modifications alone were able to predict the relationship between hormone concentration and expression level for all three iSynTFs at all expression levels. Using our expanded Hill model fit to multidimensional experimental data (hormone dose response at multiple expression levels) we were able to computationally explore the full design space of two-step transcriptional cascades, totalling 16 possible circuits. We experimentally cross-validated these simulations for a subset of circuits, confirming the predictive power of the model. These results serve as an example of the type of multidimensional biological part characterization that is required to predict circuit behavior as a composition of component parts.

RESULTS AND DISCUSSION

To predict the behavior of iSynTFs in genetic circuits, we attempted to fit a simple Hill model to the hormone dose response of each iSynTF in isolation. We first studied GEM, a previously described iSynTF that consists of the Gal4 DBD, estrogen HR, and Msn2 AD, which activates transcription from the pGAL1 promoter in response to estradiol (E2).[13] We constitutively expressed GEM from pRNR2, a medium strength constitutive promoter previously characterized in the yeast toolkit (YTK),[27] and measured its dose response: the fluorescence output of pGAL1:yellow fluorescent protein (YFP) as a function of E2. A simple Hill model accurately reproduced the basal activity, output saturation, and curvature of this pRNR2:GEM dose response (Figure 1A; see Methods).

The output of GEM is dependent on its hormone input, but this hormone dose response relationship may be modulated in nontrivial ways by the expression level of GEM itself. This effect could become significant if GEM is used in a circuit in which its expression level changes. We therefore next sought to understand the relationship between GEM expression level and its hormone dose response. Using the simple Hill model fit to the pRNR2:GEM data, we simulated the dose response of GEM at multiple expression levels around pRNR2 (Figure 1B). Changing the GEM expression level (represented by X in the simple Hill model) simply changed the half-max point of the sigmoidal hormone dose response curve, while maintaining the same basal activity, output saturation, and curvature.

We experimentally tested this prediction by measuring the dose response of GEM at several different expression levels using promoters of different strengths picked from the YTK part library.[27] We selected two promoters, pREV1 and pTEF1, that have lower and higher expression levels than pRNR2, and

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

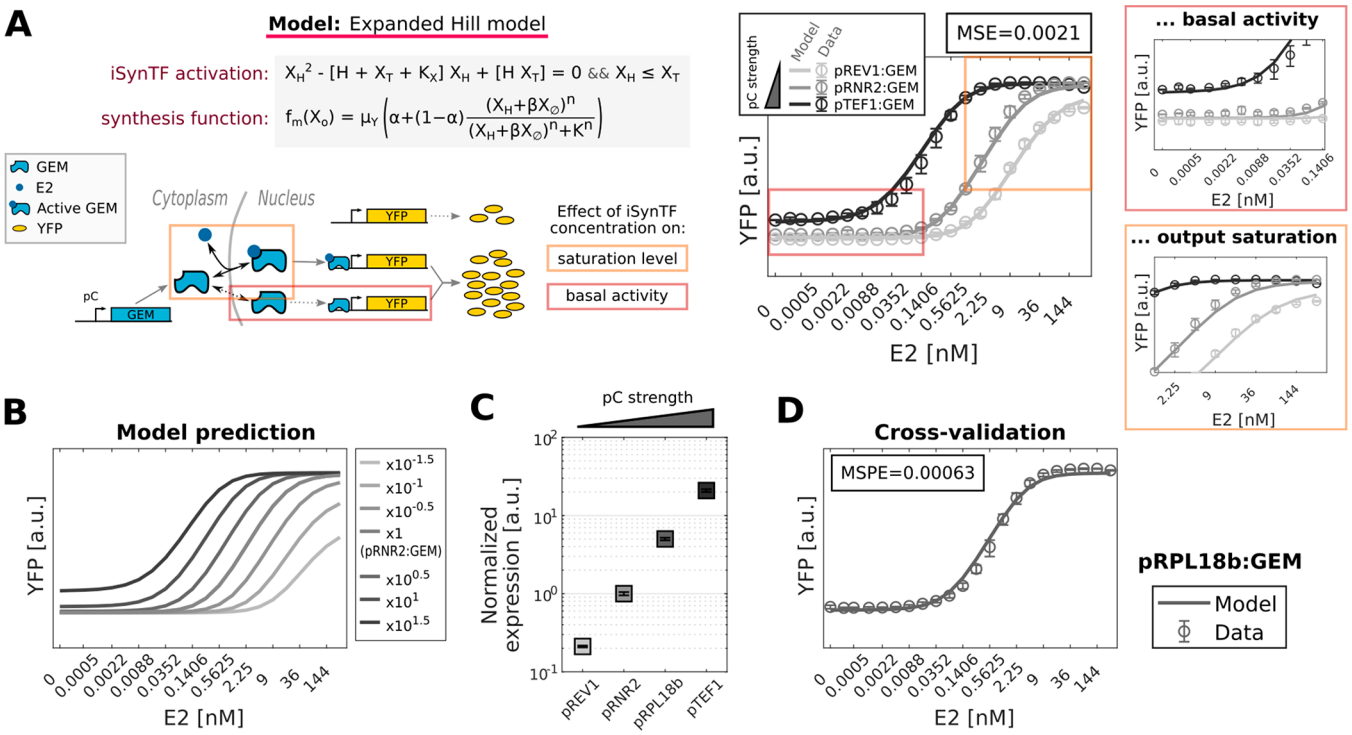


Figure 2. Expanded Hill model fit to hormone dose responses at multiple iSynTF expression levels enables accurate prediction of part behavior. (A) Left, Schematic of an expanded Hill model where the iSynTF activation step is explicitly considered (X_H and $X_\varnothing$, active and inactive iSynTF concentration; H, total hormone concentration; K_X, hormone dissociation constant; μ_Y, maximum synthesis rate; α and β, basal activity level of the free promoter and the inactive iSynTF; K, activation coefficient; n, Hill coefficient; see Methods for details). Right, Inducer (estradiol, E2) dose response of GEM at three expression levels (pREV1, pRNR2, and pTEF1 constitutive promoters). The expanded Hill model (described in the left) was fit to the observed data (mean squared error, MSE = 0.0021). Insets in red and orange boxes highlight the recapitulation of the basal activity and output saturation (compare to Figure 1D). (B) Expanded Hill model prediction of inducer dose responses for different expression levels of GEM (see legend for fold-change values). (C) Measurement of constitutive promoter expression levels using a pC:YFP fusion including pRPL18B (where pC represents one of pREV1, pRNR2, pRPL18B, or pTEF1). (D) Comparison of model prediction and experimental data for pRPL18B:GEM inducer dose response as cross-validation. Solid lines represent model predictions, open circles and filled squares represent experimental mean, and error bars represent s.d. of three biological replicates. See Supplementary Figure S1 for the equivalent analysis using the Z3PM and Z4EM iSynTFs. See Supplementary Table S4 for used parameter values.

confirmed their relative expression levels using a promoter fusion to YFP (Figure 1C). We then used these promoters to drive expression of GEM and experimentally measured each hormone dose response. Contrary to the prediction of the simple Hill model, changing the expression level of GEM did not just shift the hormone dose response half-max point (Figure 1D). We also observed a direct effect of GEM expression level on the basal activity (Figure 1D, red highlight), the output saturation (Figure 1D, orange highlight), and the curvature of each hormone dose response. These results demonstrate that iSynTFs are dose responsive in two dimensions: hormone concentration and iSynTF expression level. This prompted us to re-examine the choice of model and data used to fit the model.

We hypothesized that the Hill model predictions failed in part because this simple model did not have sufficient resolution to describe the nonlinear effect of expression level on iSynTF behavior. Furthermore, we hypothesized that we used insufficient data to fit the original model. To address the former, we introduced two new parameters to our model: one to account for the affinity between the hormone and iSynTF, and a second to account for the basal activity of the iSynTF in the absence of hormone[28] (Figure 2A; see Methods). We refer to this augmented model as the expanded Hill model. To address the latter, we fit this expanded Hill model with the hormone dose responses of GEM at three different expression levels: pREV1:GEM, pRNR2:GEM, and pTEF1:GEM. This new model was able to recapitulate all three of the experimental hormone dose responses (Figure 2A), and predicted a clear relationship between GEM expression level and the basal activity, output saturation, and curvature (Figure 2B).

To test the accuracy of the expanded Hill model, we selected a constitutive promoter of intermediate expression level from the Yeast Toolkit (YTK) part library, pRPL18B, to drive expression of GEM. We measured the pRPL18B expression level relative to pREV1, pRNR2, and pTEF1 via a YFP promoter fusion (Figure 2C) and input this information into the expanded Hill model to predict the dose response of pRPL18B:GEM. Gratifyingly, we found that the model accurately reproduced the basal activity, output saturation, and curvature of the experimental pRPL18B:GEM hormone dose response on which it was not trained (Figure 2D).

To generalize these results beyond GEM, we examined two other iSynTFs: Z3PM (a fusion of the Zif268 DBD, progesterone HR, and Msn2 AD) and Z4EM (a fusion of the Z4 synthetic zinc finger DBD, estrogen HR, and Msn2 AD).[17] Z3PM activates transcription from pZ3 in a dose responsive fashion to progesterone (Pg, Figure S1A), and Z4EM activates transcription from pZ4 in a dose responsive

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

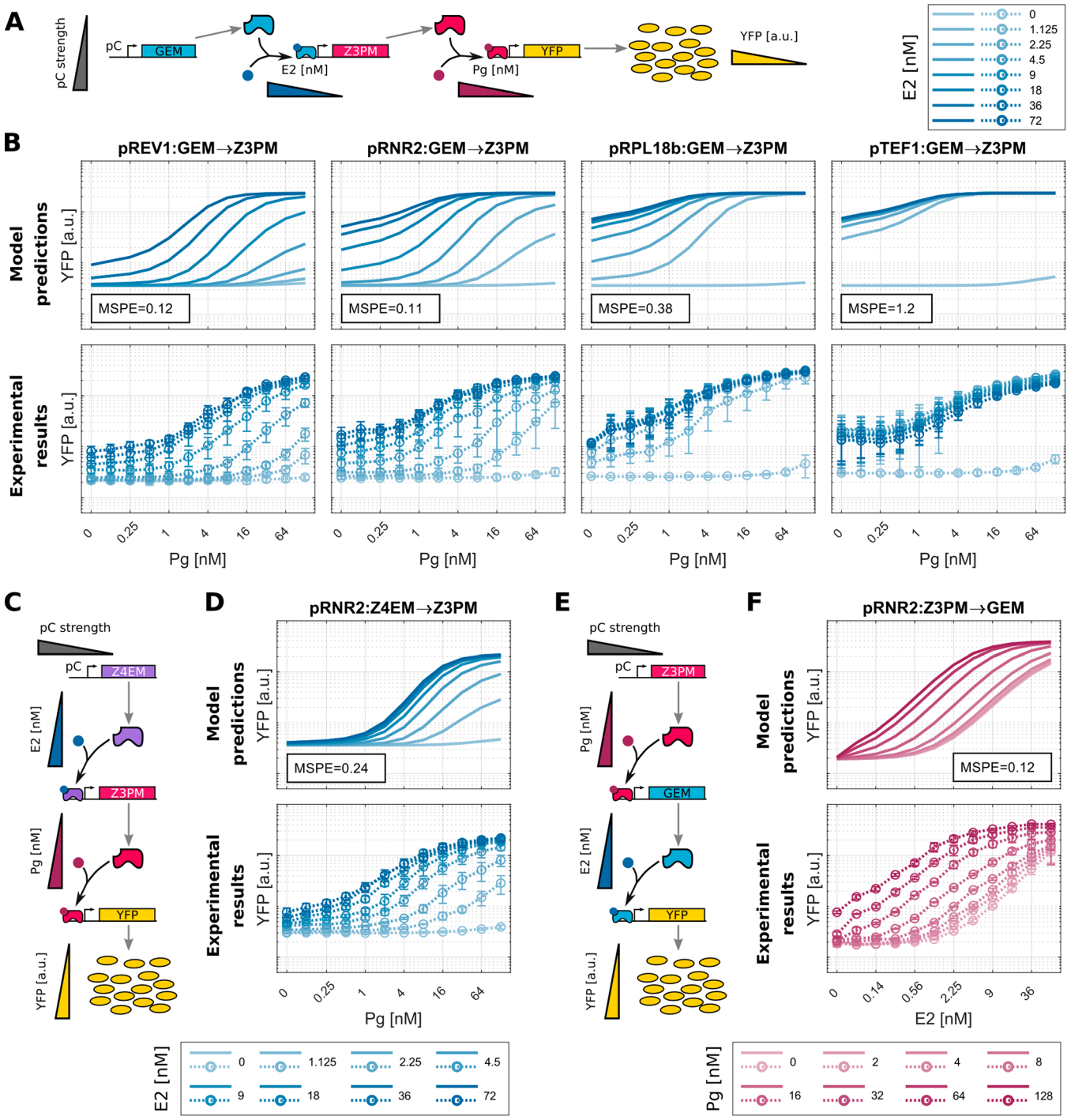


Figure 3. Using refined models to explore circuit designs. (A,C,E) A constitutively expressed (pC, constitutive promoter) iSynTF ((A) GEM, (C) Z4EM, and (E) Z3PM) is bound by its hormone inducer ((A,C) estradiol, E2, and (E) progesterone, Pg) and activates transcription of a second downstream iSynTF ((A,C) Z3PM, and (E) GEM), which in turn binds its hormone inducer ((A,C) progesterone, Pg, and (D) estradiol, E2) and activates transcription of the downstream YFP reporter. (B,D,F) Comparison of the expanded Hill model predictions (top panels) and experimental data (bottom panels) for the circuits described in (B) panel A, (D) panel C, and (F) panel E. (B) YFP expression as a function of progesterone (*x*-axis, Pg) at four different expression levels of the GEM (see plot titles); the progesterone dose responses were simulated or measured at eight different estradiol concentrations (E2, see legend in the top). (D) YFP expression as a function of progesterone (*x*-axis, Pg) at one expression level of the Z4EM (pRNR2:Z4EM), and eight different estradiol concentrations (E2, see legend in the bottom). (F) YFP expression as a function of estradiol (*x*-axis, E2) at one expression level of the Z3PM (pRNR2:Z3PM), and eight different progesterone concentrations (Pg, see legend in the bottom). The mean squared prediction error (MSPE) is shown for each case as an inset box. Solid lines represent model predictions, open circles and dotted lines represent experimental mean, and error bars represent s.d. of three biological replicates. See Supplementary Figure S3 for simulations of all circuit designs. See Supplementary Table S4 for used parameter values.

fashion to estradiol (E2, Figure S1B). To characterize these iSynTFs, we repeated the workflow developed for GEM: we expressed Z3PM and Z4EM from pREV1, pRNR2, and pTEF1, experimentally measured each hormone dose

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

response, and used these data to fit specific parameters for each iSynTF to the same expanded Hill model as above. Using the fit models, we next simulated the effect of iSynTF concentration on the hormone dose response. Simulations of Z3PM and Z4EM displayed similar trends to GEM, but they showed a much greater effect of iSynTF expression level on the basal activity and output saturation. Lastly, we cross-validated the accuracy of the Z3PM and Z4EM models against the pRPL18B expression level dose response (Figure S1C−E). The Z4EM model accurately captured the basal activity, output saturation, and shape of the pRPL18B dose response curve. The Z3PM model reproduced the output saturation, but it underestimated the basal activity and overestimated the sharpness of the curve.

When comparing the model fittings for each iSynTF, we found that multiple parameter sets fit the observed data equally well, producing similar dose response profiles (Figure S2). For the quantile 10 (*i.e.*, lower 10%) of 1000 fitting chains (see Methods), the best fitting parameter sets for each iSynTF showed that most kinetic parameter values were very well constrained (*e.g.*, basal activity α), while some appeared underdetermined (*e.g.*, hormone:iSynTF affinity constant, K_X, for GEM). It may be possible to further resolve differences between the simulations and experiments with a more detailed description of the hormone regulation. Depending on the specific application in mind, the amount and type of inaccuracy that can be tolerated from a model and its prediction varies. Here, we were most interested in fitting the qualitative trends of the hormone dose responses as a function of iSynTF expression level, and thus proceeded to use these fittings to predict genetic circuit behavior.

Using expanded Hill models of GEM, Z3PM, and Z4EM fit to multidimensional data, we explored all possible variants of a two-step transcriptional cascade: a circuit configuration where the constitutively expressed first iSynTF induces expression of a second iSynTF, which in turn induces expression of a YFP reporter (Figure 3A; see Methods). With two orthogonal HRs, there are four possible configurations of the three iSynTFs (GEM → Z3PM, Z4EM → Z3PM, Z3PM → GEM, Z3PM → Z4EM). Taking into account four possible expression levels for the first iSynTF (pREV1, pRNR2, pRPL18B, pTEF1), in total there are 16 possible circuit variants. Because GEM, Z3PM, and Z4EM each have a unique response to hormone and changing expression level, we expected that each circuit variant would behave differently in response to the two hormone inducers. In agreement, the simulations displayed different responses to both inducers, basal activities, and output saturations (Figure S3). These multidimensional fit models enable efficient screening of these circuit variants, guiding the selection of designs to be tested experimentally.

We sought to verify the accuracy of the model simulations by experimentally measuring the hormone dose responses of a subset of circuit variants. First, we studied the effect of changing the first iSynTF expression level for a single configuration by measuring the output of GEM → Z3PM at all four expression levels of GEM (Figure 3B). We found that the model accurately predicted several key aspects of the circuit behavior such as the changing output saturation and curvature. However, the simulations underestimated the effect of E2 on the basal activity in the absence of Pg.

We next investigated whether both the expanded Hill model and multidimensional data are required to predict the behavior of the pC:GEM → Z3PM circuit. First, we assessed whether the simple Hill model fit to either a single promoter dose response data or multidimensional data could predict pC:GEM → Z3PM circuit behavior (Figure S4). When we fit the simple Hill model to a single promoter dose response, the circuit behavior predictions varied significantly based on the promoter used for fitting (Figure S4A−C). As expected, most of these predictions failed to capture the cascade circuit behavior. We attempted to improve the predictions of the simple Hill model by fitting using multidimensional data (Figure S4D), but found that the simple Hill model is incapable of accounting for the basal activity and output saturation observed in our experimental data. Therefore, fitting individual parts using multidimensional data is not sufficient to predict circuit behavior if the model lacks the complexity to match the observed data.

We also observed that the expanded Hill model fit to a single promoter pRNR2:iSynTF dose response (Figure S5A) failed to properly predict the response of the pC:GEM → Z3PM circuit. Using a pair of weak promoters, pREV1:iSynTF and pRNR2:iSynTF (Figure S5B), to fit the dose response data for the iSynTF did not improve the prediction accuracy of the circuit, likely because this set of data does not fully constrain the behavior of either iSynTF. When we picked a different pair of promoters, pREV1:iSynTF and pTEF1:iSynTF (Figure S5C), which capture the low and high end of possible iSynTF expression level, the accuracy of the model prediction was very similar to the expanded Hill model fit to either three (Figure 3B) or four promoters (Figure S5D). This result shows that multidimensional data used for fitting must be carefully selected to capture the full range of responses for the parts described. Taken together, this comparison highlights that both the multidimensional characterization of the iSynTFs and the expanded Hill model are necessary to predict circuit behavior.

We used mean squared prediction error (MSPE, see Methods for definition) as a metric to compare the performance of our various models and fitting strategies. The MSPE values for the expanded Hill model fit to pREV1 and pTEF1 versus the same model fit to three or four promoters were very similar, validating our qualitative observations. However, these MSPE values were very similar to the MSPE values for the simple Hill model fit to pRNR2 data, despite the inability of the simple Hill model to capture the basal activity, output saturation, and shape of the dose response curves. This suggests that using metrics such as MSPE alone to compare models or fitting strategies may not provide a fair assessment under all contexts.

Next, we compared three circuit configurations (GEM → Z3PM, Figure 3B, Z4EM → Z3PM, Figure 3D, Z3PM → GEM, Figure 3F) at the pRNR2 expression level of the first iSynTF. As predicted by the model simulations, the Z3PM → GEM configuration displayed the greatest responsiveness to the second TF inducer in the absence of the first TF inducer. The simulations were also able to qualitatively predict the curvature of the second TF inducer dose response curves, as well as the effect of the first TF inducer on output saturation. As before, the simulations underestimated the effect of the first TF inducer on circuit output in the absence of the second TF inducer.

This slight quantitative discrepancy might be explained by a shortcoming in the model's ability to predict the expression level of the second TF as a function of the first TF inducer. The model assumes that expression of the second TF will be

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

equivalent to expression of a YFP reporter, despite the fact that contextual factors such as transcript length (*e.g.*, YFP vs iSynTF), 5′ UTR, or terminator sequence (*e.g.*, tPGK1 vs tSSA1) have a known effect on output.[29−31] Despite these shortcomings, the simulations were still able to predict key qualitative aspects of the experimental data based on the circuit configuration and expression level. Taken together, these data indicate that models can serve as a guide to genetic circuit design when an appropriate characterization of individual parts is performed.

Characterization of biological parts can be greatly simplified using technology such as cell-free systems, which remove complications of host context. Indeed, the use of cell-free extracts have proven valuable for rational, model-guided design of complex circuits in these systems.[32−34] While there has been progress toward developing models to guide circuit design *in vivo*,[15,35] obtaining quantitative precision with predictions remains challenging. Quantitative model fits can be important in certain scenarios, such as building models to automate genetic circuit design. Recently an algorithm was developed to automate the design of genetic logic gates given a set of user constraints and a library of transcriptional repressors.[11,12,36] The algorithm was successful at designing most circuits, but was not perfect; failed circuits adopted intermediate states that did not meet the digital threshold as a result of unexpected part behavior. This issue of unpredictable part behavior plagues synthetic biology in general and is a thorn in the side of many modeling efforts. Ribozymes[37] can insulate circuits from contextual effects, but inevitably there are nontranscriptional factors that may impact circuit performance. One solution is to explicitly model these factors, such as circuit−host interaction,[38,39] that may alter the predicted circuit behavior. Another powerful solution is the use of -omics level approaches such as RNA-seq or Ribo-seq to gain highly detailed information about the functionality of each circuit component.[40,41] This information can be used to rationally debug a circuit and may aid in parametrizing more complex models of circuits.

An alternative application of part models, given the universal issue of unpredictable part behavior, is theory-guided exploration of potential circuit behavior. Theory can reveal circuit topologies that produce a desired phenotype, and has been used in the past to study biochemical adaptation.[42,43] However, insights gained from these studies can often be difficult to translate into actual designs because there is no guarantee that biological parts exist in the required parameter regimes to implement such circuit designs. It may be possible to use our part characterization methodology to constrain the parameter space of theoretical explorations, biasing the results toward circuits that can be constructed using existing parts. However, our results suggest that parts would need to be characterized based on the design goal of the circuit. For example, dynamic part data would need to be collected if dynamic circuit behavior is desired, and the functionality of parts under stressors such as glucose depletion may be important if the circuit is expected to function under stress-inducing conditions.

Model based simulation of genetic circuit behavior can guide circuit designs and limit the number of constructs that need to be tested to achieve a desired behavior. The basis of accurate circuit models are accurate part models. In this work, we explored both the amount of model complexity and experimental data required to achieve qualitative prediction of circuit behavior. We found that optimization of either the model or fitting data alone were insufficient to generate accurate predictions. Our results highlight the necessity of modeling biological part behavior in the functional context of potential circuit designs and carefully designing experiments to enable parametrization of these models. The multidimensional characterization we detail in this work is a step toward careful, systematic evaluation of failure modes in circuit models and identifies one possible method for overcoming these challenges. While such an approach may require an upfront investment of time, it can pay dividends in the long term by shortening the design-build-test cycle for more complex circuits. The rational engineering of biological circuits will depend on serious efforts to devise system identification methods that can provide predictive computational representations of biological building blocks.

METHODS

Construction of DNA Constructs. Hierarchical golden gate assembly was used to assemble plasmids for yeast strain construction.[27] Individual parts were ordered as gBlocks (IDT) or PCR amplified (NEB Q5 High-Fidelity 2x Master Mix). PCR products were purified with a GeneJET PCR Purification Kit (Thermo Fisher Scientific). These sequences were domesticated with FastDigest Esp3I (Thermo Fisher Scientific). Transcriptional cassettes were constructed using BsaI-HF v2 (NEB). Multigene plasmids were constructed using FastDigest Esp3I. Plasmids are listed in Supplementary Table S1, and oligos are listed in Supplementary Table S2. The pZ4 sequence was modified from McIssac *et al.*[17] to remove a Gal4 binding site (sequence in Supporting Information).

Chloramphenicol and Ampicillin resistant plasmids were transformed into chemically competent Mach1 *E. coli* (QB3Macrolab), while Kanamycin resistant plasmids were transformed into chemically competent XL1 Blue *E. coli* (QB3Macrolab). Cultures were grown over the course of the day (Mach1) or overnight (XL1) before prep. Following growth, cultures were prepared using a GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Part plasmids were verified by sequencing (Elim Biopharmaceuticals) using the listed sequencing primers, while all other plasmids were verified by restriction enzyme digestion.

Yeast Growth Media. Overnight yeast cultures were grown in YPD (1% w/v bacto-yeast extract; 2% w/v bacto-peptone; and 2% w/v dextrose). Yeast transformation cultures were diluted into fresh YPD. Cultures for flow cytometry were diluted into SDC (0.67% w/v Difco yeast nitrogen base without amino acids; 0.2% complete supplement mixture (MP Biomedicals); and 2% w/v dextrose). For prototrophic selection following yeast transformation, SDC agar plates with the appropriate selection were used (Teknova).

Construction of Yeast Strains. All DNA constructs were transformed into a yeast strain derived from BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) that had the HIS3 locus repaired. Yeast transformations were performed as described previously[27] with modifications. One wash with 100 nM lithium acetate was performed. DNA was combined with 115 μL of transformation mixture and incubated at 42 °C for 30 min. All DNA constructs were genomically integrated. Three microliters of prepared plasmid were linearized for integration in a 20 microliter NotI-HF (NEB) reaction for 1 h and then added to the transformation mixture without purification. Strains are listed in Supplementary Table S3.

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Flow Cytometry Experiments. Yeast strains were streaked out onto YPD plates from glycerol stocks. Individual colonies were picked into 1 mL of YPD in a 2 mL V-bottom 96-well block (Corning/Costar) for overnight growth at 30 °C and 900 rpm in a Multitron shaker (Infors HT). For the individual iSynTFs experiments, overnight cultures were diluted 1:500 in 12 mL of fresh SDC in an 8-row block and 450 μL were aliquoted into a row across 2 new 96 well blocks. For the cascade experiments, overnight cultures were diluted 1:500 in 45 mL of fresh SDC in a 50 mL trough (Corning) and 400 μL were aliquoted into all wells of a new 96 well block. The YFP-promoter fusion strains were diluted 1:500 in 500 μL of fresh SDC in a new 96 well block. Following dilution, blocks were returned to the shaker for a 2 h outgrowth.

During the 2 h outgrowth, estradiol (Sigma-Aldrich) and progesterone (Fisher Scientific) induction gradients were prepared by one-to-one serial dilution. All gradients were prepared as 10× concentrated solutions.

For the individual iSynTF experiments, 50 μL of the corresponding solution were added to the 450 μL of culture, such that final, maximum concentrations of 288 nM and 256 nM of estradiol and progesterone, respectively, were used. For the cascade characterizations, 50 μL of each solution were added to the 400 μL of culture in each well in the corresponding combinations, such that final, maximum concentrations of 72 nM and 128 nM of estradiol and progesterone, respectively, were used. Blocks were then returned to the shaker for 4 h.

Following the 4 h induction, the cultures were prepared for flow cytometry. One hundred microliters of culture were mixed with 100 μL of fresh SDC in a 96-well U-bottom microplate (Greiner Bio-One). Samples were measured on a BD LSRFortessa X20 (BD Biosciences) using a high-throughput sampler. YFP-Venus fluorescence was measured using the FITC-H channel (voltage = 434). Measurements were normalized by dividing by SSC-H (voltage = 200). Analysis was performed with Python 3.7, custom scripts, and the FlowCytometryTools package. All experiments were performed in triplicate, with replicates collected on separate days. Reported values represent the mean and standard deviation of median normalized fluorescence values of the triplicates.

Model: Simple Hill Function. Under this model, we assume that the iSynTF is constitutively produced and has reached its steady state concentration (X). The concentration of hormone in the media is denoted by H. Then the steady-state concentration of the reporter protein (Y) is described as a simple Hill function with maximum synthesis rate μ_Y, basal activity $\alpha \in [0, 1]$, dissociation constant K, Hill coefficient n, and degradation rate γ:

$$f_{sh}(X, H) = \mu_Y\left(\alpha + (1-\alpha)\frac{(X\cdot H)^n}{(X\cdot H)^n + K^n}\right) \quad (1)$$

$$Y = f_{sh}(X, H)/\gamma \quad (2)$$

Model: Expanded Hill Function. As shown in Figure 1, the simple Hill model described above fails to capture the effect of the iSynTF concentration (X) on the basal expression level and saturation of the inducer dose response output (Y). To better recapitulate the observed behavior, we include the allosteric regulation of the iSynTF by the inducer in our model with the following considerations:

1. In the absence of hormone ($H = 0$), increasing iSynTF expression (*e.g.*, using a stronger constitutive promoter) increases the output expression level, suggesting some leaky or basal activation of the regulated promoter by free iSynTF.
2. As the iSynTF expression decreases, a minimum output expression level is observed for low hormone concentrations, suggesting some leaky expression of the regulated promoter independent of the iSynTF.
3. As both iSynTF and hormone concentration increase, the output expression level saturates at a maximum value.
4. With low iSynTF expression, the output expression level saturates at a lower level as hormone concentration increases, suggesting that the stoichiometric relationship of the iSynTF and hormone molecules might play an important role in output regulation.

Then, the proposed model is

$$0 = X_H^2 - (H + X + K_X)X_H + HX \quad \text{such that } X_H \le X \quad (3)$$

$$f_Y(X_H, X) = \mu_Y\left(\alpha + (1-\alpha)\frac{(X_H + \beta(X - X_H))^n}{(X_H + \beta(X - X_H))^n + K^n}\right) \quad (4)$$

$$Y = f_Y(X_H, X)/\gamma \quad (5)$$

Eq 3 represents the allosteric regulation of iSynTF by the hormone, where X_H is the active iSynTF (*i.e.*, bound to the hormone, H), and X is the total iSynTF concentration in the cell (*e.g.*, determined by the used promoter driving iSynTF expression). H is the total intracellular hormone concentration, which is assumed constant throughout the experiment, and proportional to the amount added to the media. K_X is the dissociation constant associated with the hormone-iSynTF interaction. The synthesis rate of the regulated promoter $f_Y(X_H, X)$ is still modeled as Hill function but dependent on the active iSynTF and a fraction (β) of the inactive iSynTF in the nucleus ($X - X_H$), with Hill coefficient n and dissociation constant K. As in the simple Hill model, μ_Y is the maximum synthesis rate given the translocation rate and gene, and $\alpha \in [0, 1]$ is the basal expression of the output gene (in the absence of iSynTF). Then the output steady state is simply the synthesis rate function ($f_Y(X_H, X)$) over the output degradation rate (γ; eq 5). Here, (1) the parameter β represents the basal activation by free iSynTF; (2) the parameter α represents the leakiness of the regulated promoter; (3) eq 4 considers the possible saturation with large hormone and iSynTF; and (4) eq 3 incorporates the stoichiometric relationship between the free iSynTF, hormone, and active iSynTF. This model recapitulates most of the qualitative behavior of the iSynTF regulation for several constitutive promoter strengths and hormone concentrations (see Figure 2 and Figure S1).

Model: Two-Step Transcriptional Cascade. We explore the circuit design of a two-step transcriptional cascade, where the constitutively expressed first iSynTF (X_1) induces expression of a second iSynTF (X_2), which in turn induces expression of a YFP reporter (Y). Each of these iSynTFs is modeled using the expanded Hill model (eqs 3–5; Figure 3).

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

First we solve the steady state of the second iSynTF (X_2) as a function of the first iSynTF concentration (X_1) and its inducer (H_1): $X = f_{Y,1}(X_{1,H}, X_1)/\gamma$, using the parameters fit to the first iSynTF regulation. Then, we calculate the YFP steady state (Y) by simply feeding the second iSynTF concentration and its inducer (H_2) to the model again: $Y = f_{Y,2}(X_{2,H}, X_2)/\gamma$, using the parameters fit to the second iSynTF regulation.

Model: Fitting. The goal is to minimize the error between the observed data (D) and the model prediction (Y) for a given model and parameter set (θ). We define our error function simply as the sum of squared errors in logarithmic scale:

$$\chi^2 = \sum\left(\frac{\Sigma(\log_{10}(D) - \log_{10}(Y))^2}{2\sigma_{\log_{10}(D)}}\right)$$

The mean squared error (MSE) and mean squared prediction error (MSPE) metrics shown in the figures correspond to the error function described above normalized by the number of data points considered, against either the data used for fitting the model (MSE) or new data to be predicted (MSPE).

We use a metropolis random walk (MRW) algorithm to explore the parameter space implemented as follows:

1. Choose some initial parameters θ_1 and calculate its fitting error $\chi^2_{(1)}$.
2. Iterate over $t = \{1, 2, ..., t_{max}\}$ as follows:
   a. Draw a random proposal $\phi \sim \theta_{(t)} \times 2^N\|\theta\|^{(0,\Sigma)}$ where $N_{\|\theta\|}(0, \Sigma)$ is a multivariate normal distribution with the same dimension as $\theta_{(t)}$, mean zero, and covariance matrix $\Sigma = 0.1$.
   b. We construct a likelihood function using a Gaussian function:

   $$P(D|\theta) = \exp(-\chi^2)$$

   where θ is the set of parameter to be optimized, D is the optimal data, and χ^2 is the error function. Note the likelihood is maximal when the error is minimal. Then we calculate the likelihood ratio: $\frac{P(D \mid \phi)}{P(D \mid \theta_{(t)})} = \exp(-\chi^2_{\phi} + \chi^2_{(t)})$

   Accept the proposed ϕ if the ratio is larger than a random number $\sim U[0, 1]$. The proposed value is always accepted if the error is smaller (*i.e.*, it is better).
   c. Update parameters $\theta_{(t+1)} \leftarrow \phi$ with probability $\min\left(1, \frac{P(D \mid \phi)}{P(D \mid \theta_{(t)})}\right)$; otherwise, $\theta_{(t+1)} \leftarrow \theta_{(t)}$.

For each model-data combination (Figures 1A, 2A, S1D,E, S4, S5), the MRW algorithm was run for 1000 chains starting from random parameter sets chosen from a log-uniform distribution, and each chain was run for 20 000 iterations (*i.e.*, t_{max} = 20 000). We enforced that the parameters stay in a realistic range with the following limits: $K_X = [1 \times 10^{-3}, 1000]$ nM; $\beta = [2 \times 10^{-7}, 0.2]$; $n = [1 \times 10^{-5}, 10]$; $K = [1 \times 10^{-4}, 100]$ nM; $\alpha = [2 \times 10^{-7}, 0.2]$; $\mu_Y = [2 \times 10^{-6}, 2]$ nM min^{-1}. We assume YFP dilution/degradation rate has a fixed value, γ_Y = 0.01 min^{-1}, and the measured fluorescence arbitrary units ([a.u.]) are proportional to the molecule concentration ([nM]), with 1 a.u.= 0.4 nM. The parameter values listed in Supplementary Table S4 correspond to the parameter set with the observed minimum MSE for all MRW. Supplementary Figure S2 shows the parameter values and model behavior for the lowest observed MSE for each MRW for the expanded Hill model fit to the inducer dose response of GEM, Z3PM, and Z4EM at three expression levels (pRNR2, pRNR2, and pTEF1 constitutive promoters).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.0c00288.

Supplemental figures; plasmids, oligos, and strains used; pZ4 (-Gal4 site) sequence (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Hana El-Samad − *Department of Biochemistry and Biophysics and Cell Design Institute, University of California, San Francisco, San Francisco, California 94158, United States; Chan−Zuckerberg Biohub, San Francisco, California 94158, United States;* orcid.org/0000-0001-6239-9916; Email: hana.el-samad@ucsf.edu

Andrew H. Ng − *Cell Design Institute and Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94158, United States;* orcid.org/0000-0001-8044-1807; Email: andrew.ng@ucsf.edu

Authors

Mariana Gómez-Schiavon − *Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94158, United States;* orcid.org/0000-0002-0955-7257

Galen Dods − *Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94158, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.0c00288

Author Contributions

#M.G.-S. and G.D. contributed equally to this work.

Author Contributions

G.D., M.G.-S., H.E.-S., and A.H.N. designed the study and experiments. G.D. and A.H.N. performed experiments. M.G.-S., H.E.-S., and A.H.N. developed the computational models. M.G.-S. performed the computational simulations. G.D., M.G.-S., and A.H.N. wrote the manuscript. G.D., M.G.-S., H.E.-S., and A.H.N. edited and approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the members of the El-Samad lab for useful comments and discussion on the manuscript. This work was supported by the Defense Advanced Research Projects Agency, Contract No. HR0011-16-2-0045 to H.E.-S. and National Science Foundation grant DBI-1548297 to H.E.-S. The content and information does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. H.E.-S. is a Chan-Zuckerberg investigator.

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

■ ABBREVIATIONS

iSynTF, inducible synthetic transcription factor; YFP, yellow fluorescent protein; GEM, Gal4 DNA binding domain, estradiol ligand binding domain, Msn2 activating domain; Z3PM, Z3 DNA binding domain, progesterone ligand binding domain, Msn2 activating domain; Z4EM, Z4 DNA binding domain, estradiol ligand binding domain, Msn2 activating domain.

■ REFERENCES

(1) Elowitz, M. B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature 403*, 335−338.

(2) Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature 403*, 339−342.

(3) Purnick, P. E. M., and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol. 10*, 410−422.

(4) Kwok, R. (2010) Five hard truths for synthetic biology. *Nature 463*, 288−290.

(5) Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems. *Biotechnol. J. 7*, 856−866.

(6) Del Vecchio, D., Ninfa, A. J., and Sontag, E. D. (2008) Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol. 4*, 161.

(7) Murphy, K. F., Balázsi, G., and Collins, J. J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. U. S. A. 104*, 12726−12731.

(8) Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2007) Accurate prediction of gene feedback circuit behavior from component properties. *Mol. Syst. Biol. 3*, 143.

(9) Stricker, J., et al. (2008) A fast, robust and tunable synthetic gene oscillator. *Nature 456*, 516−519.

(10) Davidsohn, N., et al. (2015) Accurate predictions of genetic circuit behavior from part characterization and modular composition. *ACS Synth. Biol. 4*, 673−681.

(11) Chen, Y., Zhang, S., Young, E. M., Jones, T. S., Densmore, D., and Voigt, C. A. (2020) Genetic circuit design automation for yeast. *Nat. Microbiol 5*, 1349.

(12) Nielsen, A. A. K., et al. (2016) Genetic circuit design automation. *Science 352*, No. aac7341.

(13) Aranda-Díaz, A., Mace, K., Zuleta, I., Harrigan, P., and El-Samad, H. (2017) Robust Synthetic Circuits for Two-Dimensional Control of Gene Expression in Yeast. *ACS Synth. Biol. 6*, 545−554.

(14) Mercer, A. C., Gaj, T., Sirk, S. J., Lamb, B. M., and Barbas, C. F., 3rd. (2014) Regulation of endogenous human gene expression by ligand-inducible TALE transcription factors. *ACS Synth. Biol. 3*, 723−730.

(15) Donahue, P. S., Draut, J. W., Muldoon, J. J., Edelstein, H. I., Bagheri, N., and Leonard, J. N. (2020) The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat. Commun. 11*, 779.

(16) McIsaac, R. S., et al. (2011) Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae. *Mol. Biol. Cell 22*, 4447−4459.

(17) McIsaac, R. S., Oakes, B. L., Wang, X., Dummit, K. A., Botstein, D., and Noyes, M. B. (2013) Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res. 41*, No. e57.

(18) McIsaac, R. S., Gibney, P. A., Chandran, S. S., Benjamin, K. R., and Botstein, D. (2014) Synthetic biology tools for programming gene expression without nutritional perturbations in Saccharomyces cerevisiae. *Nucleic Acids Res. 42*, No. e48.

(19) Pratt, W. B. (1998) The hsp90-based chaperone system: involvement in signal transduction from a variety of hormone and growth factor receptors. *Exp. Biol. Med. 217*, 420−434.

(20) Smith, D. F., and Toft, D. O. (2008) Minireview: The Intersection of Steroid Receptors with Molecular Chaperones: Observations and Questions. *Mol. Endocrinol. 22*, 2229−2240.

(21) Ng, A. H., et al. (2019) Modular and tunable biological feedback control using a de novo protein switch. *Nature 572*, 265−269.

(22) Langan, R. A., et al. (2019) De novo design of bioactive protein switches. *Nature 572*, 205−210.

(23) Hackett, S. R, Baltz, E. A, Coram, M., Wranik, B. J, Kim, G., Baker, A., Fan, M., Hendrickson, D. G, Berndl, M., and McIsaac, R S. (2020) Learning causal networks using inducible transcription factors and transcriptome-wide time series. *Mol. Syst. Biol. 16*, No. e9174.

(24) McIsaac, R. S., Petti, A. A., Bussemaker, H. J., and Botstein, D. (2012) Perturbation-based analysis and modeling of combinatorial regulation in the yeast sulfur assimilation pathway. *Mol. Biol. Cell 23*, 2993−3007.

(25) Del Vecchio, D., and Murray, R. M. (2014) *Biomolecular Feedback Systems*, Princeton University Press.

(26) Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2005) Gene regulation at the single-cell level. *Science 307*, 1962−1965.

(27) Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol. 4*, 975−986.

(28) Phillips, R., et al. (2019) Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. *Annu. Rev. Biophys. 48*, 121−163.

(29) Chen, Y.-J., et al. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods 10*, 659−664.

(30) Kosuri, S., et al. (2013) Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A. 110*, 14024−14029.

(31) Mutalik, V. K., et al. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods 10*, 354−360.

(32) Niederholtmeyer, H., Sun, Z. Z, Hori, Y., Yeung, E., Verpoorte, A., Murray, R. M, and Maerkl, S. J (2015) Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife 4*, No. e09771.

(33) Hu, C. Y., Takahashi, M. K., Zhang, Y., and Lucks, J. B. (2018) Engineering a Functional Small RNA Negative Autoregulation Network with Model-Guided Design. *ACS Synth. Biol. 7*, 1507−1518.

(34) Agrawal, D. K., et al. (2018) Mathematical Modeling of RNA-Based Architectures for Closed Loop Control of Gene Expression. *ACS Synth. Biol. 7*, 1219−1228.

(35) Bashor, C. J., et al. (2019) Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science 364*, 593−597.

(36) Taketani, M., et al. (2020) Genetic circuit design automation for the gut resident species Bacteroides thetaiotaomicron. *Nat. Biotechnol. 38*, 962−969.

(37) Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., and Voigt, C. A. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol. 30*, 1137−1142.

(38) Liao, C., Blanchard, A. E., and Lu, T. (2017) An integrative circuit-host modelling framework for predicting synthetic gene network behaviours. *Nature Microbiology 2*, 1658−1666.

(39) Sickle, J. J., Ni, C., Shen, D., Wang, Z., Jin, M., and Lu, T. (2020) Integrative Circuit-Host Modeling of a Genetic Switch in Varying Environments. *Sci. Rep. 10*, 8383.

(40) Gorochowski, T. E., et al. (2017) Genetic circuit characterization and debugging using RNA-seq. *Mol. Syst. Biol. 13*, 952.

(41) Gorochowski, T. E, Chelysheva, I., Eriksen, M., Nair, P., Pedersen, S., and Ignatova, Z. (2019) Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol. Syst. Biol. 15*, No. e8719.

(42) Ma, W., Trusina, A., El-Samad, H., Lim, W. A., and Tang, C. (2009) Defining network topologies that can achieve biochemical adaptation. *Cell 138*, 760−773.

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(43) Briat, C., Gupta, A., and Khammash, M. (2016) Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks. *Cell Syst 2*, 15−26.

https://dx.doi.org/10.1021/acssynbio.0c00288
ACS Synth. Biol. XXXX, XXX, XXX−XXX