ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

A Computational Method for Automated Characterization of Genetic Components

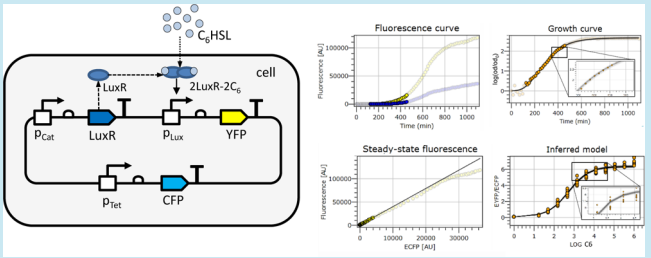
Boyan Yordanov,[†] Neil Dalchau,[†] Paul K. Grant,[†,‡] Michael Pedersen,[†,‡] Stephen Emmott,[†] Jim Haseloff,[‡] and Andrew Phillips*[,†]

[†]Microsoft Research, Cambridge CB1 2FB, U.K.

[‡]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, U.K.

Supporting Information

ABSTRACT: The ability to design and construct synthetic biological systems with predictable behavior could enable significant advances in medical treatment, agricultural sustainability, and bioenergy production. However, to reach a stage where such systems can be reliably designed from biological components, integrated experimental and computational techniques that enable robust component characterization are needed. In this paper we present a computational method for the automated characterization of genetic components. Our method exploits a recently developed multichannel experimental protocol and integrates bacterial growth modeling, Bayesian parameter estimation, and model selection, together with data processing steps that are amenable to automation. We implement the method within the Genetic Engineering of Cells modeling and design environment, which enables both characterization and design to be integrated within a common software framework. To demonstrate the application of the method, we quantitatively characterize a synthetic receiver device that responds to the 3-oxohexanoyl-homoserine lactone signal, across a range of experimental conditions.



KEYWORDS: *fluorescent reporters, ratiometric, characterization, Bayesian inference, synthetic biology*

The application of engineering principles to the design and construction of biological systems with predictable behavior is a primary focus of synthetic biology. While significant progress on the physical construction of systems[1] has been enabled by large-scale DNA synthesis and assembly methods,[2] the design of systems with predictable behavior remains a major challenge. One important design strategy has been the development of standardized parts[3] that can be composed to form devices, with the ultimate goal of assembling these devices into larger-scale systems.[4] Although this strategy has met with some success,[5,6] all but the simplest designs typically require trial and error experimentation. Furthermore, exploring these designs experimentally using high throughput methods still leads to a combinatorial explosion in the number of experiments, rendering the complete exploration of the design space infeasible even for simple systems. Computational design approaches have the potential to address this challenge through *in silico* experimentation. Such approaches require robust methods to quantitatively characterize the behavior of biological components, either as parts or as simple devices engineered for modularity,[7,8] across a range of experimental conditions. Integrated experimental and computational techniques that allow such characterization will be essential for scaling up the design process.

Fluorescent reporter proteins provide a convenient experimental tool for characterizing quantitative properties of biological components,[9] and this approach has been integrated into computational tools for the processing of fluorescence measurements.[10] To date, fluorescence-based techniques have been used to characterize a number of biological parts for the construction of synthetic systems.[5,6,11] However, the dependence of these techniques on specific experimental conditions has so far limited their broader applicability. To address this, experimental strategies based on *in vivo* measurement standards have been proposed,[12] where relative rather than absolute quantities are recorded, similar to the use of reference signals in other engineering fields. Extending this approach, spectrally distinct reporter proteins within the same device have been proposed for the measurement of relative signals, leading to the development of a ratiometric experimental protocol and an accompanying mathematical framework.[13] However, as large-scale biofabrication facilities become operational, there is an increasing need for automated, high-throughput characterization methods that integrate experimental protocols with computational analysis and are robust to changes in experimental conditions.

In this paper we present a computational method for the robust characterization of genetic components. At the core of the method, Bayesian parameter estimation and model

Special Issue: IWBDA 2013

Received: September 30, 2013

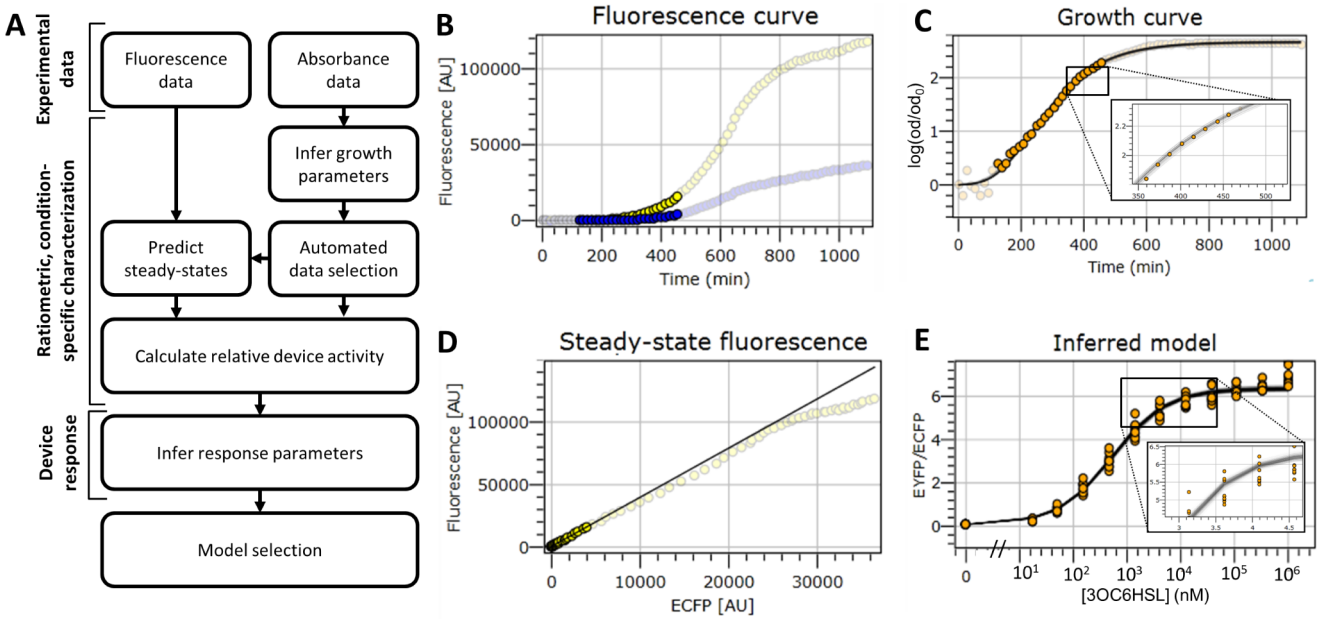


Figure 1. Overview of the computational method. (A) Processing steps of the computational method, where arrows indicate dependencies between steps. Data collection steps are performed in parallel, while data processing steps are performed sequentially. Absorbance measurements of cell growth are used to estimate a growth model (C), which also takes into account the uncertainty in model parameters (inset of C). The growth model is used to automatically select a subset of the data corresponding to the exponential growth phase (highlighted data points). The raw fluorescence measurements (B), here shown in blue (CFP) and yellow (YFP), are used to estimate the steady-state value of the ratio between the two measurements. This is obtained by computing the slope of the line that best fits the data in exponential phase (D). These estimates are used to compute the activity of the device across different experimental conditions, which here correspond to varying concentrations of chemical inducer. This allows the construction of a model describing the device response (E). The model also takes into account the uncertainty in the parameters (inset of E). Panels B–E were generated automatically from the GEC implementation of the method (see Supporting Information for details).

selection are used to evaluate competing models against experimental observations, while incorporating the uncertainty associated with the data. Our Bayesian approach can be used to determine which mechanistic hypotheses about component behavior are most likely, given the experimental data. The growth of bacteria under various experimental conditions is modeled explicitly, allowing key properties of the studied host to be inferred. Additional data processing steps allow the transcriptional activity of the device to be estimated, together with its response to external inputs such as inducer concentrations. All computational procedures are amenable to automation, which facilitates the high throughput application of the method. When combined with automated ratiometric fluorometry, the method provides an assessment of device behavior in response to changing environmental conditions, together with the resulting effect on cell growth. We implemented the method within the Genetic Engineering of Cells (GEC) modeling and design environment,[14] which provides simulation, visualization and analysis functionalities in a convenient web interface. This enables both characterization and design to be integrated within a common software framework. To demonstrate the application of the method, we quantitatively characterize a synthetic receiver device that responds to the 3-oxohexanoyl-homoserine lactone (3OC6HSL) signal, across a range of experimental conditions. We construct the device using components from a bacterial quorum sensing system,[15] for which many of the biological mechanisms have been identified.[16] The natural system has been studied both *in vivo*[17] and *in vitro*,[18] and a "datasheet" for a synthetic device based on these components has been produced.[11] Even so, diverging computational modeling strategies have been applied in previous studies using analogous synthetic devices,[19−21] where not only kinetic parameters but also mechanistic assumptions vary significantly. By characterizing the synthetic receiver using our proposed method, we provide a robust foundation for future design projects involving this component.

RESULTS AND DISCUSSION

Computational Method. The method described in this paper relies on the experimental technique developed by Brown et al.,[13] in which a characterization experiment is performed by simultaneously measuring a number of signals over time, including absorbance, proportional to the number of cells in a culture, together with the fluorescence of spectrally distinct reporter proteins. Briefly, the method relies on the use of two fluorescent reporter signals, one to monitor the transcriptional activity of a device and another to monitor the activity of a reference promoter. This allows a direct *ratiometric* measurement, given by the ratio of the device activity to the reference promoter activity. The ratio is more robust to variations in experimental conditions and instrumentation, which affect both fluorescent signals simultaneously and are thus attenuated.[13] The experimental procedure can also be used to measure the activity of a device in response to inputs such as chemical inducers.

In our computational procedure (Figure 1), the raw data is first corrected to account for the effects of background and autofluorescence (see Methods). A set of growth parameters is then inferred for each sample using the absorbance signal (the optical density OD_{600}). To characterize the device during exponential growth, a relevant subset of the data is selected automatically. This differs from the previous approach,[13] for which manual tuning was required. The selected data is then

dx.doi.org/10.1021/sb400152n | *ACS Synth. Biol.* XXXX, XXX, XXX−XXX

used to predict the absolute steady-state value of each fluorescent signal, which we refer to as F_i^s. A direct calculation is then used to obtain the relative steady-state fluorescence, which is used to compute the relative device activity. The device can be characterized under varying experimental conditions, corresponding to differences in temperature, pH, or concentration of a chemical inducer. The response of the device to changes in these conditions can be characterized by inferring the parameters of an appropriate mathematical model. Different mathematical models, representing alternative hypotheses for the mechanisms governing the device response, can also be rigorously compared. In the following we highlight the key computational steps of the method that enable its automation and implementation within the GEC framework[14] (Supplementary Figure S1).

Bayesian Parameter Inference. The characterization procedure outlined above aims to infer the parameters of mathematical models describing the behavior of a biological device. We accomplish this through Bayesian parameter inference methods, generalized to a variety of functional relationships. Formally, given a data set $D = \{(x_0, y_0), ..., (x_n, y_n)\}$ and a model $y = f(\theta, x)$ with parameters θ, let $P(\theta|D)$ denote the *posterior* distribution, which describes the joint conditional probability of different parameter combinations given the data. The posterior captures the uncertainty about the exact model parameters given uncertainty in the experimental data. This uncertainty can be propagated in subsequent computational studies involving the inferred model, allowing a better assessment of the expected device behavior as opposed to singleton (e.g., average) parameter estimates. In particular, this provides a statistically rigorous form of sensitivity analysis, which can be used to determine the robustness of predictions to uncertain measurements and parameter estimates.

Using Bayes' formula, the posterior is expressed as

$$P(\theta|D) = \frac{P(D|\theta)\cdot P(\theta)}{P(D)} \tag{1}$$

Based on this, estimates of the posterior are obtained using an MCMC procedure (see Methods) by assuming a uniform *prior* parameter distribution $P(\theta)$ and defining the *likelihood* of a parameter set as

$$P(D|\theta) = \prod_{i=0}^{n} \mathcal{N}(y_i, \hat{y}_i, g(\hat{y}_i)) \tag{2}$$

where $\mathcal{N}(y, \mu, \sigma) = 1/(\sigma(2\pi)^{1/2})e^{-[(y-\mu)^2/(2\sigma^2)]}$ denotes the normal (Gaussian) distribution with mean μ and variance σ, and $\hat{y}_i = f(\theta, x_i)$ is the model prediction for parameter set θ. Intuitively, we assume that the experimental measurements are samples drawn from the model but corrupted by Gaussian measurement noise. We compute the likelihood of each data point (x_i, y_i) using the Gaussian probability distribution centered around the ideal $\hat{y}_i$. The variance of the experimental error often depends on x_i. We refer to this relationship as a *noise model* and represent it by the function $g(x)$. One family of noise models relevant for this work is given as $g(x) = \sigma\cdot x^n$ and includes constant ($n = 0$), linear ($n = 1$), and square-root ($n = 1/2$) noise models.

Growth Models and Data Selection. While characterizing the transcriptional activity of a device is the main focus of the method, the growth of cells containing the device under different experimental conditions is also recorded. To study the effects of the device on cell growth, we infer the parameters of a growth model for each experimental sample, based on the optical density (OD) signal. Several classical growth models[22] were implemented, but in the following we focus on the Gompertz model:

$$\ln\left(\frac{\mathrm{OD}(t)}{\mathrm{OD}(0)}\right) = K\,\exp\left[-\exp\left(\frac{\mu\cdot e}{K}(\lambda - t) + 1\right)\right] \tag{3}$$

where $e = \exp(1)$ is Euler's constant, K is the carrying capacity of the medium, λ is the lag length, and μ is the maximal (exponential) growth rate. We infer the growth parameters K, λ, and μ from the experimental data using the Bayesian procedure described above. We use a constant noise mode, which assumes that the standard deviation does not scale with the output. This allows us to characterize the growth of cells engineered with a given device under different experimental conditions, such as media, inducer concentration, and temperature. While only the growth rate μ is used to compute the device activity, the additional growth parameters provide useful information about the studied system.

To select the range of data representing the exponential phase of cell growth, we estimate the time at which the maximal (exponential) growth is reached as

$$t_m = \frac{K}{e\cdot\mu} + \lambda \tag{4}$$

(see Supporting Information for details). The range of data used for all subsequent computation is then defined as $F_i(t)$ and $\mathrm{OD}(t)$ where $\lambda < t < 2\hat{t}_m - \lambda$ (here, $\hat{t}_m = t_m + \varepsilon$ is the time when the maximal growth rate is reached, offset by a factor $\varepsilon = 50$ min to capture the maturation of reporters). Essentially, this computation corresponds to selecting a time interval following the lag phase and centered around the time when the maximal growth rate is achieved, taking into account the delay caused by the maturation of the reporters.

Activity Computation. The ratiometric characterization method[13] relied on the computation of steady-state fluorescence values per cell of each individual signal as $F_i^s = (\partial F_i)/(\partial \mathrm{OD})$. A subset of the data was selected, for which a linear relationship between fluorescence and absorbance holds (e.g., the exponential growth phase). The absolute steady-state transcriptional activity of the device expressing reporter i was computed as

$$P_i = \frac{\gamma_M(\alpha + \gamma + \mu)(\gamma + \mu)}{\rho\cdot\alpha\cdot n} F_i^s \tag{5}$$

where μ is the growth rate of the bacteria and all other parameters describe the expression and maturation of the reporter, including protein degradation rate (γ), mRNA degradation rate (γ_M), translation rate (ρ), and maturation rate (α). It is assumed that the degradation rates of mature and immature fluorescent protein are similar. By designating one of the signals F_r as a reference, the relative steady-state activity[13] was computed from eq 5

$$R_i = \frac{P_i}{P_r} = \frac{\alpha_r(\alpha + \mu)}{\alpha(\alpha_r + \mu)}\cdot\frac{F_i^s}{F_r^s} \tag{6}$$

where α and α_r denote the maturation rates of the primary and reference reporters. (Several assumptions on reporter stability and the similarities between primary and reference signals[13] were involved in simplifying eq 6.)

dx.doi.org/10.1021/sb400152n | ACS Synth. Biol. XXXX, XXX, XXX−XXX

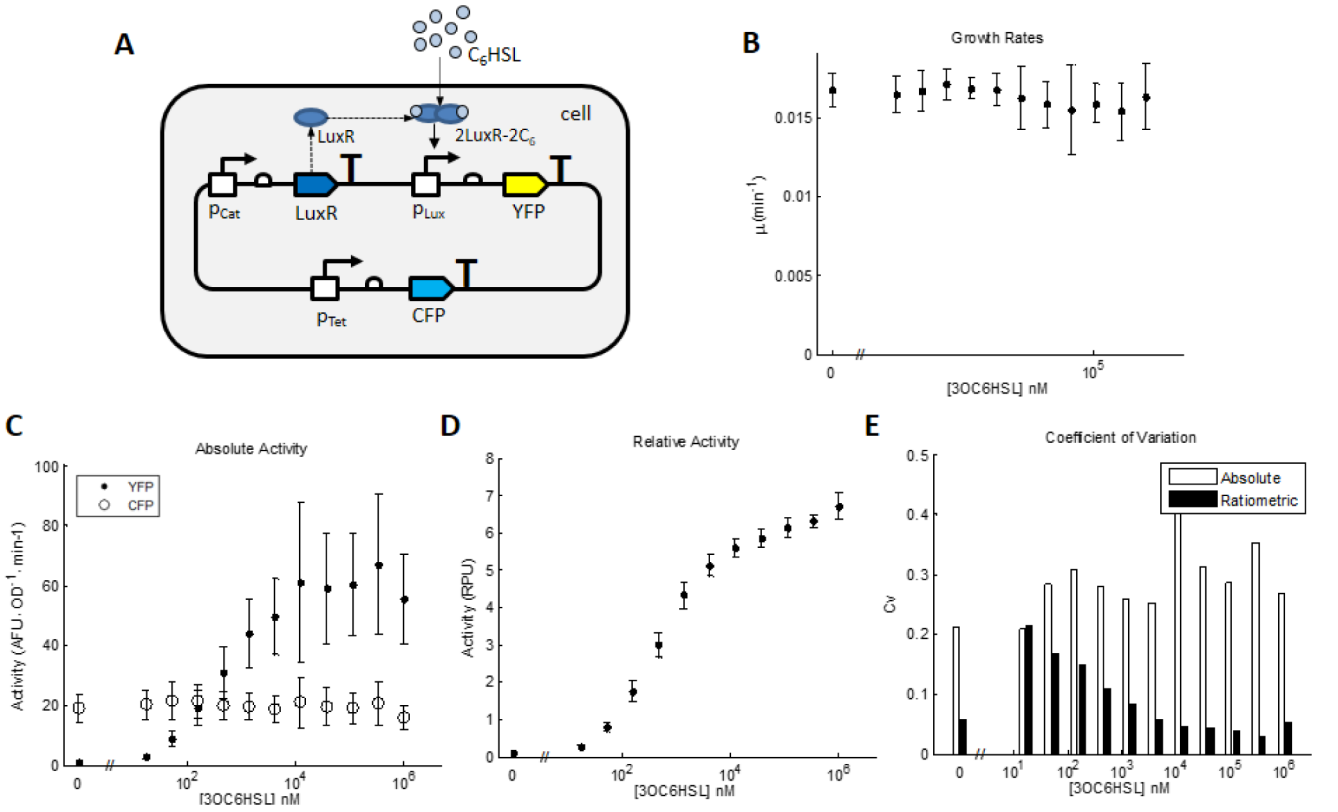


Figure 2. Quantitative characterization of a 3OC6HSL receiver device. The 3OC6HSL receiver device (A) was constructed and characterized at various concentrations of input signal. Growth rates remained unchanged across these conditions (B) and both the absolute (C) and relative (D) characterization procedures captured the induction of the device. The ratiometric strategy significantly decreased the experimental variability, leading to a more robust characterization (E). Note that the absolute measurements (C) were not adjusted for signal gain, since only the signal gain ratio between the two channels was measured (see Methods).

In our method we use eq 5 to compute the absolute steady-state device activities but implement a more direct method for the computation of relative activity as

$$R_i = \frac{\alpha_r(\alpha + \mu)}{\alpha(\alpha_r + \mu)} \cdot R_{i,\mathrm{r}} \tag{7}$$

where $R_{i,j} = \partial F_i / \partial F_j$ is a direct estimate of the steady-state fluorescence ratio per cell between a pair of signals, which is computed using linear least-squares regression. This extension allows us to eliminate the dependence on measured OD data from the computation. Furthermore, the linear relationship between each pair of fluorescent signals holds over larger ranges as opposed to the relationship between each fluorescent signal and OD. Overall, these factors contribute to a more robust computational strategy and decrease the variability between individual measurements. Note that currently only the maximum likelihood estimate of μ, rather than the full posterior, is used for the computation in eq 7.

Response Models and Model Selection. A variety of different mathematical models are applicable for describing the changes in the behavior of a device in response to changes in its environment. The choice of model depends on the specific property being studied (such as changes in transcriptional activity, growth rate, or lag phase length) and external quantity being varied (such as temperature or inducer concentration). Here, we focus on modeling the steady-state transcriptional activity of a device in response to changes in concentrations of an external regulator. For such inducible or repressible devices, the functional response

$$P = \frac{a \cdot [\mathrm{A}]^n + b \cdot K^n}{[\mathrm{A}]^n + K^n} \tag{8}$$

known as a *Hill function*[23] has been widely used for characterization. Here, P is the transcriptional activity of the device, [A] denotes the concentration of the regulator, and the remaining parameters capture the maximal (a) and basal (b) activity, sensitivity (K), and cooperativity (Hill coefficient, n). For some simple chemical mechanisms such functional response models can be derived directly. For example, this is the case when a number of inducer molecules bind a target promoter directly and the rate of expression is a from a bound promoter and b from a free promoter. Such a promoter is inducible for $a > b$ and repressible for $b < a$, with no expression from the unbound promoter when $b = 0$. The parameter n reflects the number of inducer molecules involved (provided that the binding of all n inducers is simultaneous). More generally $n > 1$, where n is not necessarily an integer, represents some degree of cooperativity between the binding of multiple regulator molecules. More complicated functional forms than the one from eq 8 have been derived for regulation mechanisms involving multiple regulatory states and regulators, leading to a generalizable framework.[24]

Once the response of a device has been characterized, it becomes possible to compare alternative mechanisms by comparing models of the expected steady-state transcriptional activity in response to changes in the experimental conditions. For this, a metric of the capacity of a model to describe the observed experimental characterization data is required, which

dx.doi.org/10.1021/sb400152n | ACS Synth. Biol. XXXX, XXX, XXX−XXX

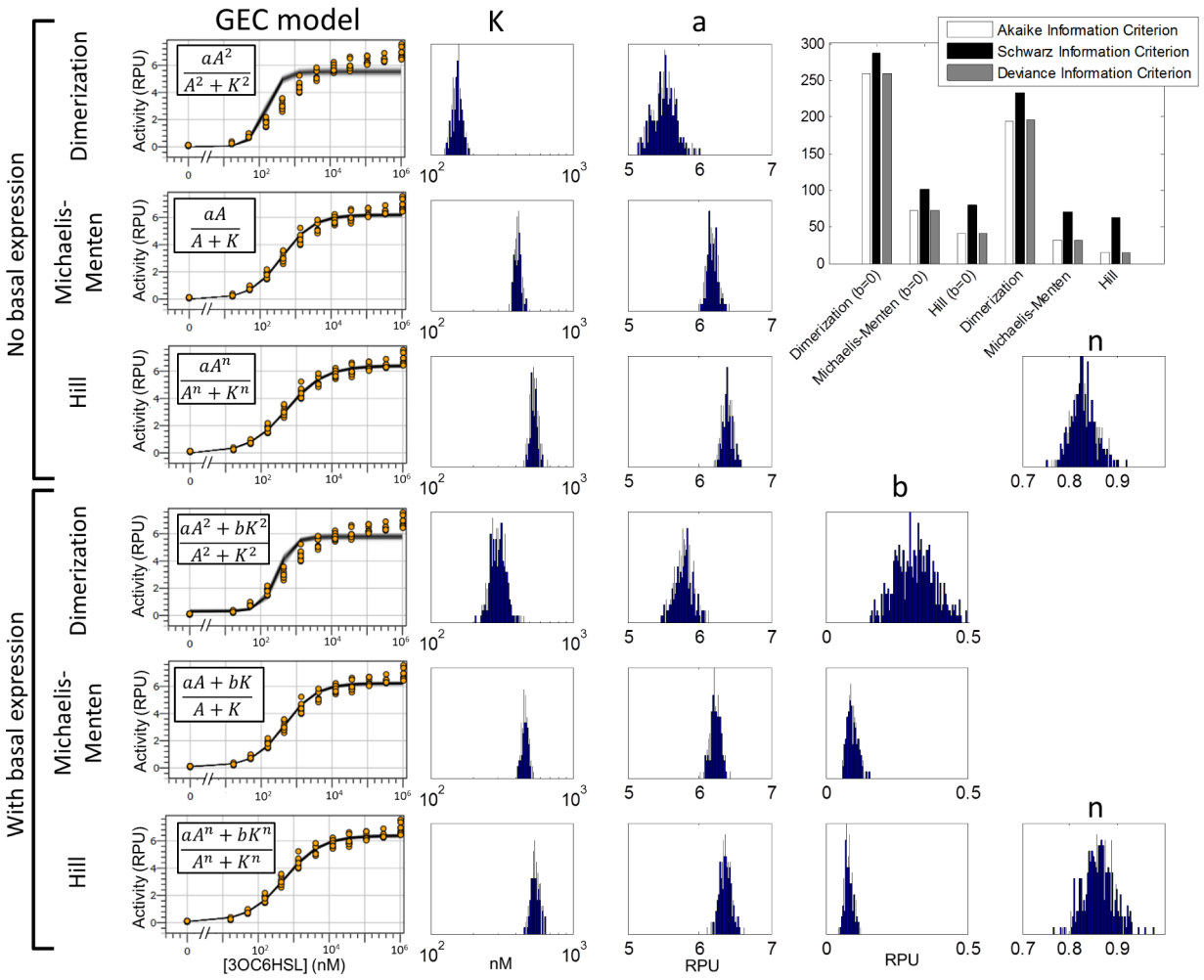


Figure 3. Model selection and parametrization for the 3OC6HSL receiver device. A number of different models corresponding to alternative mechanisms describing the 3OC6HSL receiver were compared, including dimerization (row 1), Michaelis–Menten (row 2), and Hill (row 3) mechanisms without basal expression, as well as the same mechanisms but allowing for basal expression (rows 4–6, respectively). In all cases, the GEC characterization method was used with the indicated functional relationships and the resulting model is shown (as visualized within GEC) in column 1. For each model, the histograms relate the posterior distributions (obtained from 50,000 MCMC samples after an initial burn-in phase of 50,000 samples) for parameters K (second column), a (third column), b (fourth column), and n (fifth column). Note that parameters b and n have constant values for some models where their posteriors are omitted. The information criteria indicating the capacity of these mechanisms to explain the data are compared in the top right panel. The posterior parameter distributions associated with a given model often reveal important details about the identifiability of model parameters, the suitability of parameter ranges, and whether the model is appropriate for the device being characterized.

takes into account the model complexity. This helps to protect against parameter overfitting, in which additional parameters may spuriously model exogenous variations in the data arising from experimental fluctuations, for example. For such comparisons, a number of *information criteria*, including the Akaike (AIC), Bayesian (BIC), and Deviance (DIC) information criterion are computed automatically as part of our Bayesian parameter inference procedure (see Methods). We use these metrics to assess different mathematical models describing the response of a device, which represent different mechanistic hypotheses.

Implementation and Graphical Interface. To enable both characterization and design within a common software framework, we integrated our computational method as part of the GEC environment,[14] which provides simulation, visualization, and analysis functionalities, as well as a convenient web interface. The novel characterization module provides a convenient interface for browsing the raw experimental data (see Supplementary Figure S1 for a screen shot illustrating this), as well as the computed growth, steady-state fluorescence, and activity values for each signal. A choice of several classical growth models[22] and functional models (including Michaelis–Menten, dimerization, and Hill) is available. This allows the response of any measured signal (e.g., absolute or relative activity, growth rate, carrying capacity, etc.) to experimental conditions to be studied and alternative mechanistic hypotheses describing the device behavior to be rigorously compared.

Receiver Characterization. We applied our computational methodology to the characterization of a receiver device (Figure 2), constructed using components from a bacterial quorum sensing system.[15,16] In this device, the input signal 3OC6HSL diffuses into the cell and forms a complex with the receiver protein LuxR,[17,18] which is expressed constitutively from the low expression Pcat promoter. The choice of the Pcat promoter was motivated by the negative effects on growth observed when LuxR was driven by higher activity promoters.

dx.doi.org/10.1021/sb400152n | *ACS Synth. Biol.* XXXX, XXX, XXX–XXX

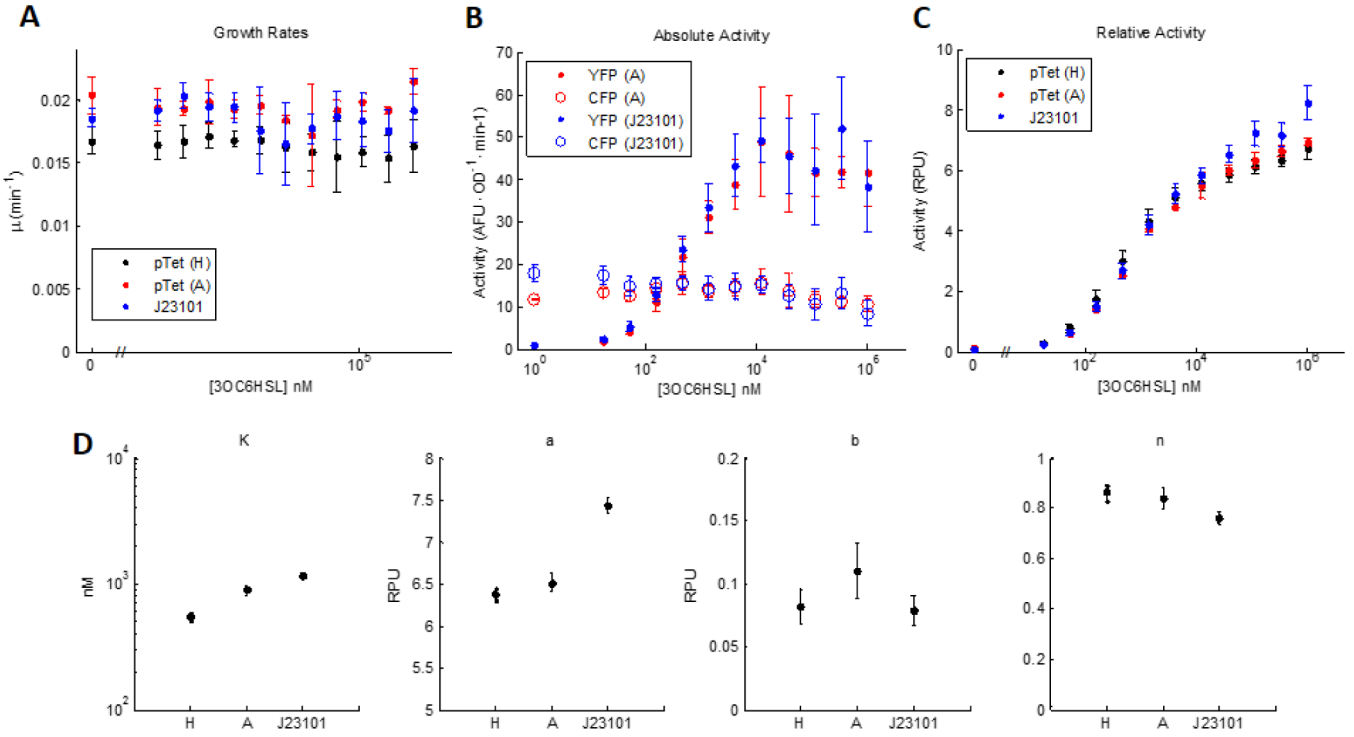


Figure 4. Robustness of the ratiometric characterization methodology. The initial characterization of the 3OC6HSL receiver (data set denoted by H and shown in black) was compared against characterization experiments performed in a different laboratory facility (data set denoted by A and shown in red) and with a different reference promoter J23101 (data set denoted by J23101 and shown in blue). The parameter values of the general Hill function (eq 8) were obtained for each data set using GEC. The posterior distributions for each parameter obtained using our procedure are available in Supplementary Figure S5, and the mean parameter values obtained for all three data sets are compared in panel D, where the error bars correspond to 1 standard deviation from the mean. Note that the relative promoter units (RPU) differ for the characterizations performed using the Ptet and J23101 reference promoters, where the RPUs for Ptet are slightly higher than for J23101. Previous experiments have characterized Ptet gainst the J23101 reference as 1.05 ± 0.08,[13] 1.2 ± 0.1,[12] and 1.29 ± 0.16 RPU.[13]

Such effects were also apparent in our LuxR induction experiments, where the receiver protein was produced from an additional high-copy plasmid (see Figure 5B). The LuxR-3OC6HSL complex subsequently binds to promoter Plux, which recruits RNA polymerase,[17,18] thus inducing the expression of the reporter signal YFP. The reference signal CFP is expressed from Ptet, which behaves as a constitutive promoter due to the absence of the TetR repressor. We also investigated the extent to which our characterization method is independent of the reference promoter used, by replacing the reference promoter Ptet with the promoter J23101.[12,13]

The behavior of the receiver device was characterized by varying the concentrations of the 3OC6HSL signal between 0 and 1 mM. Models describing the growth of bacteria under the different conditions were inferred using the Bayesian procedure described above. The results indicated that an increase in the level of induction did not significantly impact the growth rate (for example, a growth rate $\mu = 0.0173$ corresponds to a 40 min doubling time) (Figure 2A). The activity of the device was characterized by measuring the absolute YFP fluorescence (Figure 2C), together with the ratio of YFP to CFP (Figure 2D). The coefficients of variation for both measurements were compared (Figure 2E). Plots of individual experimental and technical replicates are available in Supplementary Figure S3.

To determine the functional relationship between the device activity and concentration of 3OC6HSL, we applied our parameter inference procedure to model the relationship between device activity and the concentration of 3OC6HSL. [For this experiment, separate data sets (corresponding to experimental replicates) were merged before performing parameter inference. Thus, the various corrections (gain, background, autofluorescence) are not performed individually for each experiment, but as a whole this did not affect the results significantly.] A common characterization strategy, consistent with previous studies,[13,19−21] involves fitting a Hill equation to capture the functional response (eq 8), where [A] denotes the concentration of the 3OC6HSL signal. We compared models capturing dimerization ($n = 2$), Michaelis−Menten ($n = 1$), and Hill mechanisms, both with ($b > 0$) and without ($b = 0$) basal expression, as possible mechanisms describing the receiver response.[13,19,21] The dimerization model corresponds to a mechanism where two copies of the LuxR-3OC6HSL complex bind the Plux promoter cooperatively. In contrast, the Michaelis−Menten model corresponds to the case where either a single complex binds the promoter or there is no cooperativity. Finally, the Hill model generalizes these mechanisms by introducing a cooperativity parameter n and is capable of capturing the functional response of more complicated mechanisms through fractional values of n. To correlate the studied models to the ratiometric measurements, note that the parameters a and b should be interpreted as being relative to the second channel and are thus quantified in relative promoter units (RPUs). Additional information regarding the derivation of these equations from more detailed chemical reaction mechanisms is available in the Supporting Information. We compared these different models by computing the posteriors illustrating the captured parameter uncertainties, as well as the information criteria metrics reflecting the capacity of

dx.doi.org/10.1021/sb400152n | ACS Synth. Biol. XXXX, XXX, XXX−XXX

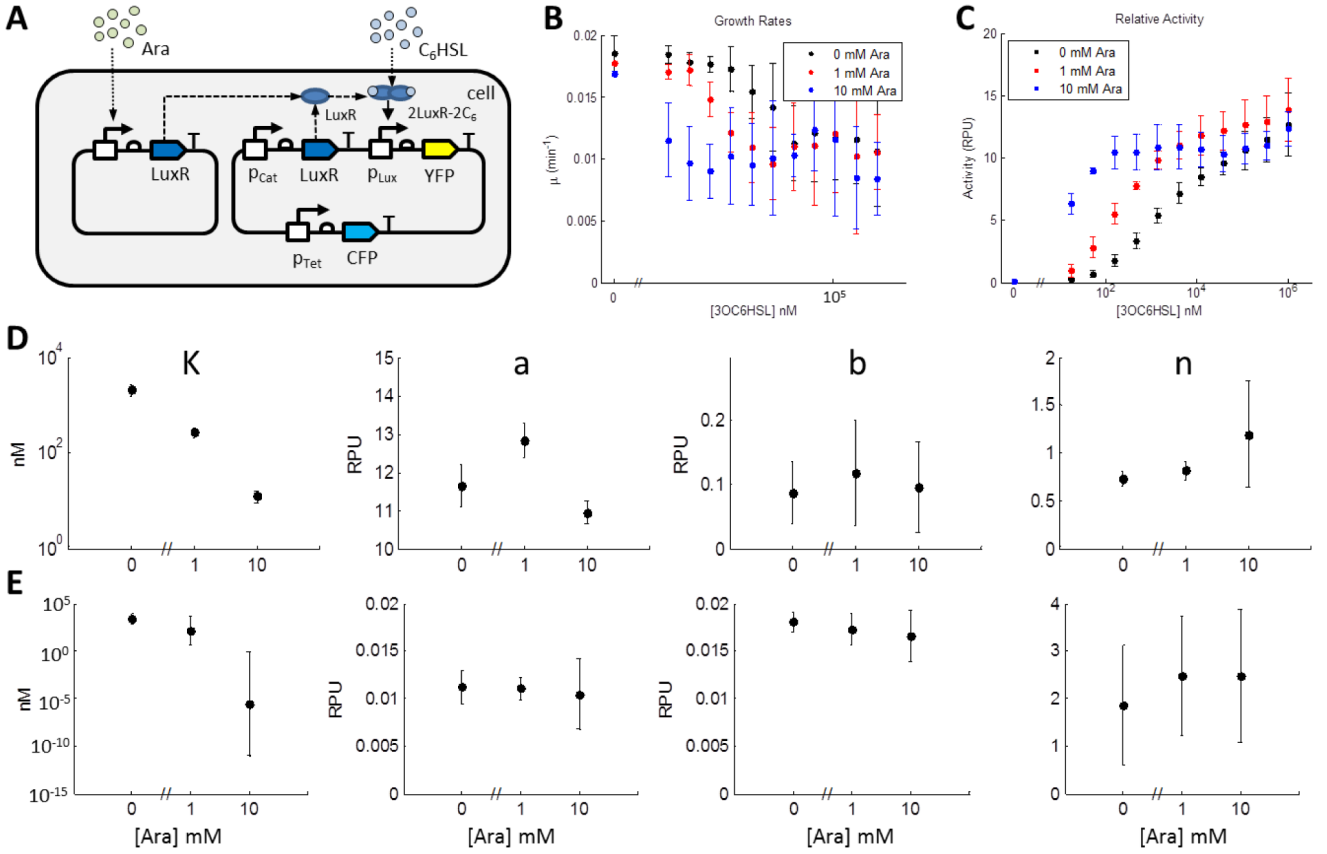


Figure 5. Enhanced 3OC6HSL sensitivity via inducible expression of LuxR. A modified 3OC6HSL receiver device (A) was constructed to allow the induction of the LuxR receiver protein through the addition of arabinose. Note that details of the additional arabinose receiver component such as the constitutively expressed AraC receiver protein are omitted from the illustration. The GEC characterization procedure was used to calculate the mean ± standard deviation of the posterior distributions for (B) growth rate and (C) relative promoter activity. The device was characterized at 0, 1, and 10 mM arabinose and at various concentrations of the 3OC6HSL signal. (D) Hill function characterization of relative promoter activity as a function of 3OC6HSL. For each concentration of arabinose, separate Hill functions were inferred against the data in panel C, with the Hill function parameters (*K*, *a*, *b*, *n*) as indicated (see Supplementary Figure S8 for full posterior distributions). (E) Hill function characterization of growth rate as a function of 3OC6HSL. For each concentration of arabinose, separate Hill functions were inferred against the data in panel B, with the Hill function parameters (*K*, *a*, *b*, *n*) as indicated (see Supplementary Figure S9 for full posterior distributions). In panels D and E, symbols and error bars represent the mean ± standard deviation of the posterior distributions, inferred via Bayesian inference in GEC.

the models to explain the data (Figure 3). Both the Hill and Michaelis−Menten mechanisms were comparable in terms of their capacity to explain the experimental data. However the more general Hill mechanism was preferred, despite the additional parameters. Including a nonzero basal expression term further improved the quality of the models. In contrast, the dimerization mechanisms (leading to a cooperative $n = 2$ term) provided a worse fit. While it is believed that LuxR-type proteins form dimers in order to interact with DNA,[16] the dimerization mechanism we studied was characterized by the cooperativity of monomer bindings. In contrast, the non-cooperative binding captured within the Michaelis−Menten and Hill models (when $n < 1$) was consistent with previous experimental observations.[18] Overall, the ratiometric characterization procedure resulted in decreased parameter uncertainty as compared to using the absolute measurements (Supplementary Figure S4).

Characterization Robustness. To investigate the robustness of the ratiometric approach, we characterized the 3OC6HSL receiver device (Figure 2A) in two additional settings. First, we characterized the device in two different laboratory facilities equipped with an equivalent plate reader. Second, we replaced the reference promoter Ptet with the promoter J23101, which has previously been used as a reference.[12,13] The results are presented in Figure 4. Note that the relative promoter units (RPU) differ between the Ptet and J23101 reference promoters. In previous experiments, the strength of the Ptet promoter (R0040) was characterized against the J23101 reference as 1.05 ± 0.08[13] and 1.2 ± 0.1 RPU[12] using absolute measurements and as 1.29 ± 0.16 RPU using ratiometric measurements.[13] This suggests that the RPUs for Ptet are slightly higher than for J23101. The response of the device to 3OC6HSL was characterized as a general Hill model using each experimental data set (Figure 4D). GEC model visualizations and full parameter posteriors are shown in Supplementary Figure S5. Parameters *a* and *b* are reported in RPUs and differ for data sets using the promoter Ptet (A and H) and the promoter J23101. The higher maximal activity *a* of the receiver when characterized against J23101 is consistent with the higher reference activity of the Ptet promoter.

Enhanced 3OC6HSL Sensitivity via Inducible Expression of LuxR. To study the effect of receiver concentration on

dx.doi.org/10.1021/sb400152n | *ACS Synth. Biol.* XXXX, XXX, XXX−XXX

device characterization, the LuxR receiver protein was expressed from the inducible Pbad promoter on a separate, high-copy plasmid (Figure 5A), in addition to the original Pcat construct. Thus, the concentration of LuxR in the system was regulated through the addition of arabinose. We characterized this device at different concentrations of 3OC6HSL (ranging from 0 to 1 mM) and at three different concentrations of arabinose (0, 1, and 10 mM) to obtain relative measurements (Figure 5C). The absolute measurements obtained for the LuxR induction experiments are available in Supplementary Figure S6. Details regarding the individual technical and experimental replicates are available in Supplementary Figure S7. As expected, changing the concentration of the receiver protein affected the sensitivity K of the device (Figure 5D), which changed by over 2 orders of magnitude ($K = 2127 \pm 566$ nM with 0 mM arabinose and $K = 12.5 \pm 3.0$ with 10 mM arabinose. The full posteriors and GEC model visualizations of the receiver response are available in Supplementary Figure S8. The results also indicated that the growth rates of bacteria carrying this device were significantly affected by induction with 3OC6HSL and, furthermore, that this effect became more pronounced upon induction with arabinose (Figure 5B). Assuming that the growth effects are related to the activity of the device, we used a Hill model to capture the response of growth to the changing experimental conditions and obtained parameters for this model (Figure 5E, Supplementary Figure S9). These results indicate that growth rates are decreased upon induction with 3OC6HSL (parameter b is greater than a in Figure 5E) and that growth rate begins to decrease at lower 3OC6HSL concentrations when more LuxR is present. Specifically, parameter K was greater for 10 mM arabinose than for 0 and 1 mM (Figure 5E).

In this paper we presented a computational method for the characterization of genetic devices, with the goal of enabling automated and robust characterization procedures. Our method was based on the use of a ratiometric experimental protocol[13] requiring the simultaneous measurement of at least two fluorescence signals, which is within the capabilities of standard laboratory equipment. Although the approach requires two distinct fluorescence reporters, this does not increase the experimental burden dramatically from single reporter methods, while enabling significantly increased accuracy of characterization. The experimental cost can be further reduced by constructing reusable characterization templates, which contain most required components with the exception of the device to be characterized.

This work provides a step toward the design and construction of reliable biological devices from characterized components. However, the application of this approach to systems of increased complexity is required to better understand component modularity. In particular, the appropriate definition of a module will need to be rigorously determined, such that the characterized properties of a device are preserved in different contexts. Here we assume that the devices being characterized are engineered for modularity through one of several approaches developed previously in the field.[7,8] Moreover, our method is general in the sense that it can be used to characterize devices at different levels of abstraction, depending on the assumptions of modularity being made.

Our method was implemented within the GEC environment, allowing the automated processing of experimental data and the parametrization of models for subsequent computational studies. The GEC implementation also improved the accessibility of the characterization method by facilitating both modeling and visualization. While a number of settings were exposed to enable interactive experimentation and tuning of the computational procedure, its automatic application is also possible using the default settings. For typical characterization experiments of individual devices, our framework supports rapid data processing and model fitting. For example, the computation times for the characterization of the receiver device were under 1 min. While currently all computation steps are performed sequentially (Figure 1), the computation for each replicate is easily parallelizable, since the same data processing steps are applied to each of the different experimental conditions. The computation of a device response, however, depends on the prior characterization of all experimental conditions and must therefore be done as a subsequent step.

By inferring a number of key parameters describing the growth of cells in various experimental conditions, we were able to gain additional insights about the effect of a synthetic device on its host. For example, we were able to quantify how a synthetic 3OC6HSL receiver device affects the growth rate of its host, for different values of input signal. Enabling these experiments motivated our use of the more restrictive functional growth models[22] rather than alternative interpolation techniques such as polynomial or spline interpolation, which have been used previously.[10] The inferred growth parameters also allowed us to automate the selection of the subset of experimental data pertaining to exponential cell growth for subsequent characterization. Measuring parts and devices during exponential growth is consistent with previous approaches,[12,13] but the characterization of components in other settings, such as during stationary phase, is also desirable. While previous approaches[13] were unable to estimate the steady-state and reference activities under these conditions, our direct steady-state ratio computation strategy is independent of the growth data. For such experiments, the behavior of the reference promoter across different growth phases is critical. For instance, characterization of a device in stationary phase will require a reference promoter that is not silenced during this phase.

Our method relied on Bayesian parameter inference to quantify the response of a device to external perturbations. We illustrated this functionality by modeling the response of the receiver to various concentrations of its input signal. The model selection capabilities of our characterization method allowed us to explore a variety of models describing this response. These models corresponded to different mechanistic hypotheses about the behavior of the device but differed in their capacity to explain the observed experimental data. This allowed us to compare several possible mechanisms. We envision that such rigorous comparisons will provide a valuable tool for exploring possible biological mechanisms, provided that robust measurement protocols leading to the collection of precise experimental data are available. The posterior distributions often reveal important information about the characterization experiment. For example, inappropriately selected parameter ranges result in truncated distributions with a sharp cutoff at the limiting value. Furthermore, the identifiability of parameters given the available data can be assessed from these distributions, which might prompt additional experiments. Finally, the uncertainty of individual parameters helps gauge the overall capacity of the model to explain the experimental data. Computing and representing this uncertainty is a major advantage of the

Bayesian parameter inference procedure over alternative strategies that produce singleton estimates (e.g., mean parameter values) or highly simplified approximations (e.g., mean parameter values with error).

To minimize variability due to experimental conditions and instrumentation, we based our method on a ratiometric characterization protocol.[13] For the characterization of the receiver device, both the absolute (Figure 2C) and relative (Figure 2D) measurements were capable of capturing the induction of the device activity, as a function of the input signal concentration. The ratiometric method significantly reduced the variability in experimental measurements, as indicated by the coefficient of variation (Figure 2E). The more precise characterization also reduced the uncertainty in the model parameters describing the response of the device to changing input signal concentrations (Supplementary Figure S4). This allowed different models, corresponding to alternative mechanistic hypotheses for the device response, to be compared more accurately (Figure 3). Out of the set of mechanisms that was investigated, the general Hill equation was best at describing the experimental data, despite involving more parameters, as indicated by the computed information criteria (top right panel of Figure 3). The noncooperative binding ($n \leq 1$) observed in Figure 3 was consistent with previously described experimental results[18] where a Hill coefficient of $n = 0.9$ was identified. The slight preference for the more-general Hill mechanism (with $n < 1$), as compared to the simplified Michaelis−Menten mechanism ($n = 1$), indicated a possibility that the true mechanism governing the behavior of this receiver involves additional interactions.

The ratiometric procedure also reduced variability when characterizing a device in different experimental settings. In our experiments, the characterization of the 3OC6HSL receiver was consistent when performed in a different laboratory setting (Figure 4), which was not surprising given the similarity of experimental protocols and laboratory instrumentation used. Even so, some differences (e.g., in the measured growth rates, Figure 4A) illustrated the variability in experimental measurement that can occur. The receiver characterization was also consistent when two different promoters, Ptet and J23101, were used as references (Figure 4). The ability to compare characterization experiments performed against different references is beneficial for certain situations, such as when the reference promoter is affected by the chemical inducer used to study the device response. Even so, comparing ratiometric measurements collected against different references is not always straightforward. While we were able to consistently characterize the receiver device against both the Ptet and J23101 reference promoters, the results were not comparable when a CI repressible promoter was used as a reference, due to interferences that are currently being investigated. Note that the reported RPU units between the relative measurements in Figure 4C were not directly comparable, as they represented the relative strength of the receiver against different reference promoters. To compare these units directly, a precise characterization of the relative strength between the two reference promoters is required. In previous experiments the strength of the Ptet promoter (R0040) was characterized against the J23101 reference as 1.05 ± 0.08 RPU[13] and 1.2 ± 0.1 RPU[12] using absolute measurements and as 1.29 ± 0.16 RPU using ratiometric measurements.[13] This is consistent with our results when parameters a and b (Figure 4D) identified for the J23101 construct are rescaled to convert the units from J23101 to Ptet RPU. In spite of the agreement, this indicates the challenges in characterizing parts and devices across different standards, where small errors in the estimation of the relative standard strengths accumulate.

Our computational method also enabled us to study the effects of environmental changes on cell growth. We observed that the growth rates of bacteria carrying the modified 3OC6HSL receiver (Figure 5A) were significantly affected by induction with 3OC6HSL (Figure 5B), as compared to the original device (Figure 2A) where growth rates remained unchanged (Figure 2B). The decrease in growth was possibly a result of the higher availability of the LuxR receiver protein, leading to higher activity of the device from Figure 5A (e.g., comparing parameters a in Figure 4D and Figure 5D). The decrease became more pronounced upon induction with arabinose (Figure 5B). These effects were also captured in the models of the growth rate response to changes in arabinose and 3OC6HSL (Figure 5E). Changing the concentration of the receiver protein LuxR also affected the sensitivity of the device (Figure 5C), which is consistent with previously described mechanisms of the 3OC6HSL, LuxR and Plux promoter binding.[25] The change in the device sensitivity was also observed in the identified response models (parameter K in Figure 5D), where the high sensitivity for the LuxR induction at 10 mM Ara resulted in the increased uncertainty in the inferred parameter n, since only a single data point was available at low 3OC6HSL induction for this experiment.

Although previous work has pursued the characterization of parts,[12] here we focus on the characterization of devices as simple reusable components in order to avoid context-specific sequence effects. However, modularity is not always guaranteed even at this level of abstraction.

Our approach focused on characterizing the steady-state activity of devices from population (plate reader) measurements, since such experimental protocols are often more accessible and scale more readily. In addition, the ratiometric protocol[13] could offer significant advantages for single-cell (e.g., flow cytometry) measurements. Extending our computational procedures for such measurements is a promising direction for future work. Notably, when devices with more complex dynamic behavior such as bistability are characterized, the heterogeneity in a population could hinder the use of population averaged signals for characterization. Similarly, while the steady-state behavior of a device provides an accurate description for some systems, the transient dynamic behavior is crucial for the correct functionality of others and must also be studied.

By providing sophisticated data processing and computational modeling capabilities within an interactive software tool, our approach will help to streamline the characterization of devices. Currently, the method is most useful for automating the computational aspects of characterization experiments, while also providing a rigorous treatment of model uncertainty and facilitating interactive exploration of model hypotheses. In future, the method will be more closely integrated with the biological design process, for example, to allow the automated characterization of biological devices within a centralized database whenever new experimental data is uploaded. This, in turn, will enable the accurate simulation and analysis of systems composed of characterized components, prior to their physical construction. Additional future extensions to the computational method and its implementation involve the ability to seamlessly handle multiple characterization data sets,

tracking several input and output signals, and allowing improved customization of growth and response models, in addition to adapting the approach to a range of organisms. We envision that further development of the approach will enable the establishment of new standards for robust characterization within the broader scientific community.

METHODS

Plasmids. Plasmids (listed in Supplementary Table S3) were constructed using Gibson assembly.[2] Ratiometric reporters were constructed by modifying J69594:pSB3K3CY[13] as described in tab:Plasmids. All constructs initiated translation at ribosome binding site (RBS) B0034 from the Registry of Standard Biological Parts (Registry) and terminated transcription with terminator B0010 (Registry).

Plate Fluorometer Assays. Plasmids described were used to transform *E. coli* strain DH10B (Ecloni 10G, Lucigen). Single colonies were used to inoculate cultures in M9 medium supplemented with 0.2% casamino acids, 0.4% glucose, and 1 mM thiamine with appropriate selective antibiotics (50 μg/mL kanamycin, 50 μg/mL carbenicillin). After overnight growth at 37 °C, these cultures were diluted 1:1000 in the same medium and transferred to 96-well clear-bottom plates (Greiner). 3-Oxohexanoyl-homoserine lactone (Sigma) was dissolved in DMSO to a concentration of 200 mM and then diluted in growth medium to the appropriate concentrations such that 5 μL of solution added to 195 μL of cell suspension resulted in the final induction concentrations described. Plates were placed in a BMG Fluostar Omega platereader set to 37 °C with shaking at 200 rpm. Optical density (600 nm) and fluorescence (eCFP: 430 nm excitation, 480 nm emission; eYFP: 500 nm excitation, 530 nm emission) readings were taken approximately every 10 min for 17 h.

Controls. Several controls were measured in order to calibrate the characterization data. Measurements of blank media was performed to calculate the median background absorbance and fluorescence of all channels, and these values were subtracted from all characterization measurements. The fluorescence of nonfluorescent cells was also measured to correct the effects of autofluorescence where linear regression was used to establish a relationship between absorbance and autofluorescence during the exponential growth phase of bacteria. Direct measurements of devices where the same promoter is driving the expression of both the eYFP and eCFP reporters were used to measure the signal gain ratios between the two channels. To obtain more accurate estimates, the J23101 promoter and the PLas promoter induced by low (0 μM) and high (25 μM) concentrations of the 3OC12HSL input signal were measured and averaged. Note that these measurements captured only the gain ratio between the eYFP and eCFP channels rather than the gains of each channel individually. All controls were measured separately for the two experimental setups used in this study.

Computation. All data processing, parameter inference, and characterization were performed using an implementation of the computational method described in this paper as an extension to the GEC tool.[14] Matlab was used to combine separate results and produce figures, but all computational components required for characterization experiments are contained within the GEC extension. For the Bayesian parameter inference used within the characterization method, Markov Chain Monte Carlo (MCMC) techniques implemented within the Filzbach tool (http://research.microsoft.com/filzbach) are employed, which also computes the information criteria used for model comparison. For the inference of growth parameters, a linear noise model was used with 2,000 burn-in and sampling steps. For the response models, a square-root noise model was used with 50,000 burn-in and sampling steps. Literature values describing the translation, degradation, and maturation rates of reporter proteins were used for the computation of device activity and are available in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: andrew.phillips@microsoft.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Jim Ajioka for laboratory facilities used to conduct comparative experiments. P.K.G. is supported by the John Templeton Foundation Grant ID 15619: Mind, Mechanism and Mathematics: Turing Centenary Research Project.

REFERENCES

(1) Gibson, D. G., et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science 329*, 52−6.

(2) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343−5.

(3) Endy, D. (2005) Foundations for engineering biology. *Nature 438*, 449−53.

(4) Purnick, P. E. M., and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol. 10*, 410−22.

(5) 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.

(6) Ellis, T., Wang, X., and Collins, J. J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol. 27*, 465−71.

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

(8) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods 10*, 354−60.

(9) de Jong, H., Ranquet, C., Ropers, D., Pinel, C., and Geiselmann, J. (2010) Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst. Biol. 4*, 55.

(10) Boyer, F., Besson, B., Baptist, G., Izard, J., Pinel, C., Ropers, D., Geiselmann, J., and de Jong, H. (2010) WellReader: a MATLAB program for the analysis of fluorescence and luminescence reporter gene data. *Bioinformatics 26*, 1262−3.

(11) Canton, B., Labno, A., and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol. 26*, 787−93.

(12) Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D.,

and Endy, D. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng. 3*, 4.
(13) Brown, J. R. (2013) A design framework for self-organised Turing patterns in microbial populations. Ph.D. thesis, Department of Plant Sciences, University of Cambridge.
(14) Pedersen, M., and Phillips, A. (2009) Towards programming languages for genetic engineering of living cells. *J. R. Soc. Interface 6*, S437−S450.
(15) Lazdunski, A. M., Ventre, I., and Sturgis, J. N. (2004) Regulatory circuits and communication in Gram-negative bacteria. *Nat. Rev. Microbiol. 2*, 581−92.
(16) Nasser, W., and Reverchon, S. (2007) New insights into the regulatory mechanisms of the LuxR family of quorum sensing regulators. *Anal. Bioanal. Chem. 387*, 381−90.
(17) Yan, L., Allen, M. S., Simpson, M. L., Sayler, G. S., and Cox, C. D. (2007) Direct quantification of N-(3-oxo-hexanoyl)-L-homoserine lactone in culture supernatant using a whole-cell bioreporter. *J. Microbiol. Methods 68*, 40−5.
(18) Urbanowski, M. L., Lostroh, C. P., and Greenberg, E. P. (2004) Reversible acyl-homoserine lactone binding to purified *Vibrio fischeri* LuxR Protein. *J. Bacteriol. 186*, 631−637.
(19) McMillen, D., Kopell, N., Hasty, J., and Collins, J. J. (2002) Synchronizing genetic relaxation oscillators by intercell signaling. *Proc. Natl. Acad. Sci. 99*, 679−684.
(20) Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., and Weiss, R. (2005) A synthetic multicellular system for programmed pattern formation. *Nature 434*, 1130−4.
(21) Song, H., and You, L. (2006) Evolving sensitivity. *ACS Chem. Biol. 1*, 681−2.
(22) Zwietering, M. H., Jongenburger, I., Rombouts, F. M., and van't Riet, K. (1990) Modeling of the bacterial growth curve. *Appl. Environ. Microbiol. 56*, 1875−81.
(23) Weiss, J. N. (1997) The Hill equation revisited: uses and misuses. *FASEB J. 11*, 835−841.
(24) Gutierrez, P. S., Monteoliva, D., and Diambra, L. (2012) Cooperative binding of transcription factors promotes bimodal gene expression response. *PloS One 7*, e44812.
(25) Haseltine, E. L., and Arnold, F. H. (2008) Implications of rewiring bacterial quorum sensing. *Appl. Environ. Microbiol. 74*, 437−45.